

**UNIVERSIDAD CENTROCCIDENTAL
"LISANDRO ALVARADO"
DECANATO DE CIENCIAS Y TECNOLOGÍA
DEPARTAMENTO DE POSTGRADO**

**ESTIMACIÓN DE MODELOS MACROECONÓMICOS
DINÁMICOS VÍA APROXIMACIONES ESTOCÁSTICAS**

Barquisimeto, Venezuela.

Julio de 2012

UNIVERSIDAD CENTROCCIDENTAL
"LISANDRO ALVARADO"
DECANATO DE CIENCIAS Y TECNOLOGÍA
DEPARTAMENTO DE POSTGRADO

ESTIMACIÓN DE MODELOS MACROECONÓMICOS
DINÁMICOS VÍA APROXIMACIONES ESTOCÁSTICAS

TRABAJO ESPECIAL DE GRADO PRESENTADO POR

LCDA. JESSICA P. PÉREZ

COMO REQUISITO FINAL
PARA OBTENER EL TÍTULO DE MSc
EN MATEMÁTICA MENCIÓN OPTIMIZACIÓN
ÁREA DE CONOCIMIENTO: ESTADÍSTICA, OPTIMIZACIÓN
TUTOR: LCDO. MSc. JHONNY ESCALONA

Barquisimeto, Venezuela.

Julio de 2012

“ESTIMACIÓN DE MODELOS MACROECONÓMICOS DINÁMICOS VÍA APROXIMACIONES ESTOCÁSTICAS”

RESUMEN

En este trabajo presentaremos un algoritmo tipo Metropolis- Hasting que usa aproximaciones estocásticas para la estimación de la moda a posteriori en una clase de modelos macroeconómicos dinámicos no lineales, denominados en la literatura como Modelos Estocásticos de Equilibrio General Dinámico (DSGE, sus siglas en inglés). Usaremos data real y data simulada sobre estos algoritmos para estudiar su eficiencia. Elaboramos experimentaciones computacionales para comparar las soluciones obtenidas con el algoritmo diseñado y un algoritmo Metropolis- Hasting.

Gracias a Dios, a la Virgen, a mi hijo, a mi familia y a todos los que creyeron que esta meta sería posible.

AGRADECIMIENTOS

A mi Dios todopoderoso por ser mi padre fiel y amigo, que con su amor incondicional me brinda el privilegio de guiarme en cada uno de mis pasos.

A mi Divina Pastora que desde el día de mi nacimiento me lleva en su regazo, gracias por guiarme para lograr mis metas.

A mi hijo por ser el tesoro más grande que tengo en mi vida, por ti lucho y trato de ser mejor cada día, eres mi razón de vivir.

A mi mamá, que es un regalo maravilloso de Dios, gracias por ser mi fortaleza, mi guía, mi amiga. Mami otra meta lograda.

A mi papá Mario gracias por su apoyo y compañía por ser uno de los pilares de nuestra familia.

A mi hermanita, Jesenia por entenderme en todo momento, por su compañía, por apoyarme y ser partícipe de este gran sueño.

A mis amigos y muy especial a Marisela por estar siempre a mi lado brindándome todo tu apoyo y amor de hermana.

A ti mi amor, por estar en mi vida brindarme tu apoyo y compañía. Hemos compartido cosas muy importante, hoy compartimos un sueño más de muchos que vendrán.

A todas aquellas personas que de una u otra forma comparten esta nueva meta de mi vida y que colaboraron conmigo en este nuevo camino que hoy toma cuerpo y se convierte en una realidad.

Gracias.

ÍNDICE

Agradecimientos	i
Introducción	1
1. Método de Aproximaciones Estocásticas	3
1.1. Teorema de Convergencia	5
2. Técnica de Aproximación Estocástica	9
2.1. Algoritmo de Aproximaciones Estocásticas	10
2.2. Estimador, $\hat{s}(\cdot)$	12
2.3. Sesgo del estimador $\hat{s}(\cdot)$	13
2.3.1. Teorema (Cota del sesgo del estimador)	13
2.4. Modificación de segundo orden del algoritmo de SPSA	15
2.4.1. Aplicación f_k para 2SPSA	17
3. Modelo Macroeconómico	20
3.1. Condiciones de Optimalidad	22
3.1.1. Estado Estacionario	23
3.1.2. Log- Linealización del Modelo	25
3.1.3. Resolución del modelo	27
3.2. Modelo Básico de Crecimiento en forma de Espacio-Estado.	29
3.3. Filtro de Kalman	30
3.3.1. Propiedades: Filtro de Kalman	33
3.4. Estimación de Máxima Verosimilitud	35
3.4.1. Función de Verosimilitud asociada al Modelo	37
3.5. Algoritmo Metropolis-Hasting	38

4. Resultados Numéricos	41
4.1. Parámetros del Modelo Macroeconómico	42
4.2. Data para la experimentación	44
4.2.1. Experimentaciones con Data Calibrada	44
4.2.2. Experimentaciones con la data real	47
Conclusiones	54
Referencias	55

ÍNDICE DE TABLAS

- 4.1. Estimaciones: MH-SPSA versus Metropolis- Hasting con la Data Calibrada 1 45
- 4.2. Estimaciones: MH-SPSA vs Metropolis- Hasting con la Data calibrada 2 46
- 4.3. Estimaciones: MH-SPSA versus Metropolis- Hasting considerando data real 47

ÍNDICE DE FIGURAS

1.1. Función con y sin Ruido	4
4.1. Estimaciones: MH-SPSA vs Metropolis- Hasting con la Data Calibrada 1	48
4.2. Estimaciones: MH-SPSA vs Metropolis- Hasting con la Data Calibrada 1	49
4.3. Estimaciones: MH-SPSA vs Metropolis- Hasting con la Data Calibrada 2	50
4.4. Estimaciones: MH-SPSA vs Metropolis- Hasting con la Data Calibrada 2	51
4.5. Estimaciones: MH-SPSA vs Metropolis- Hasting considerando Data Real	52
4.6. Estimaciones: MH-SPSA vs Metropolis- Hasting considerando Data Real	53

INTRODUCCIÓN

Para poder caracterizar el comportamiento de los diversos fenómenos como el crecimiento económico, los ciclos económicos, la política monetaria y la política fiscal se plantean los modelos macroeconómicos no lineales, conocidos en la literatura como modelos estocásticos de equilibrio general dinámico (DSGE, sus siglas en inglés). Estos modelos han tomado una gran importancia en la actualidad. El enfoque que siguen la mayoría de los autores es la estimación basada en la resolución de una aproximación lineal del problema original que luego es utilizada para obtener una representación espacio estado lineal del modelo y de esta manera se obtienen evaluaciones de la función de verosimilitud, la cual puede ser utilizada para obtener estimaciones de los parámetros a través del enfoque clásico de máxima verosimilitud y/o también puede ser utilizado el enfoque Bayesiano. Para el enfoque clásico, se suele utilizar algoritmos de optimización determinística como los algoritmos Casi-Newton, pero tales algoritmos necesitan de evaluaciones del gradiente y hessiano lo cual suele ser muy costoso computacionalmente si la dimensión del vector de parámetros a estimar es muy grande, para el enfoque bayesiano, la estimación de la distribución a posteriori del modelo se realiza con el algoritmo Metropolis-Hasting [10]. En este trabajo implementamos una combinación del algoritmo Metropolis-Hasting con un algoritmo de optimización estocástica que sólo utiliza dos evaluaciones de la función de verosimilitud para aproximar el gradiente, este algoritmo acelera la convergencia del Metropolis-Hasting [1].

La forma en que esta estructurado este trabajo es la siguiente: Comenzaremos en el Capítulo 1 con el estudio del método de aproximaciones estocásticas; en el capítulo 2 mostraremos la técnica de aproximación estocástica para la estimación de parámetros denominada SPSA, la cual puede ser utilizada para obtener los estimadores de

máxima verosimilitud, así como una modificación de segundo orden del mismo. El capítulo 3 es una descripción del problema Macroeconómico que será estudiado y llevado a su forma espacio estado pasando por el estudio de la log linealización del modelo, también describimos el Filtro de Kalman y el Algoritmo de Metropolis- Hastings.

En el capítulo 4 se muestra la descripción y análisis numérico de la estimación de máxima verosimilitud del modelo, algunas características importantes referentes a la comparación de los métodos en cuanto a convergencia y a la estimación generada. Finalmente daremos un análisis de estos resultados numéricos y las conclusiones generales.

CAPÍTULO 1

MÉTODO DE APROXIMACIONES ESTOCÁSTICAS

Los algoritmos de aproximaciones estocásticas introducidos por Robbins- Monro en 1951 [6]. Proveen de un simple y efectivo mecanismo para encontrar las raíces o mínimos de una función, cuyas evaluaciones u observaciones están posiblemente contaminadas por ruidos aleatorios.

Consideraremos el siguiente problema: dada una función $M : \mathbb{R} \mapsto \mathbb{R}$ diferenciable y α un valor constante, tal que la ecuación

$$M(x) = \alpha, \tag{1.1}$$

tiene una raíz única $x = \theta$. Existen muchos métodos para la determinación del valor de θ por aproximaciones sucesivas, por ejemplo, el método de Newton- Raphson, el cual se puede describir como un método iterativo, que comienza con una estimación inicial de la solución x_0 y se construye una sucesión de aproximaciones de forma recurrente mediante la ecuación:

$$x_{j+1} = x_j - \frac{M(x_j)}{M'(x_j)}.$$

Vamos a considerar la generalización estocástica del problema anterior, en el cual la naturaleza de la función $M(\cdot)$ es desconocida en el experimento y monótonamente creciente, por ejemplo, la función dada por $M_\epsilon(x) = M(x) + \epsilon$ (ver Figura 1.1), es una función contaminada con un cierto ruido, ϵ . Supongamos que cada valor de x corresponde a la variable aleatoria $Y = Y(x)$ con función de distribución $P_r[Y(x) \leq y] = H(y|x)$, tal que

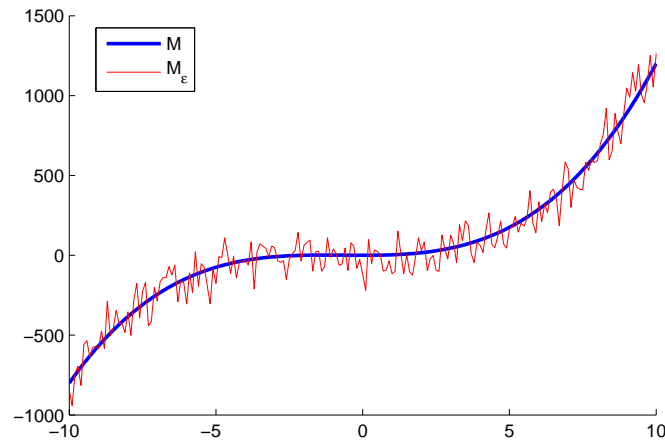


FIGURA 1.1: Función con y sin Ruido

$$M(x) = \int_{-\infty}^{+\infty} y dH(y|x), \quad (1.2)$$

es el valor esperado de Y dado x . Ni la naturaleza exacta de $H(y|x)$ ni $M(x)$ son conocida en el experimento, pero suponemos que la ecuación (1.1) tiene una única raíz, θ , y se desea estimar dicha raíz haciendo observaciones sucesivas sobre Y en los niveles x_1, x_2, \dots determinados sucesivamente en concordancia con algún procedimiento definido. Si la sucesión generada converge en probabilidad a la raíz y es independiente de cualesquiera valores iniciales arbitrarios x_1, x_2, \dots, x_r , en conformidad con la terminología estadística se puede llamar al procedimiento consistente para $H(y|x)$ y el valor de α .

En lo que sigue presentaremos un procedimiento particular dado por Robbins Monro para estimar θ , el cual es consistente bajo ciertas restricciones en la naturaleza de $H(y|x)$. Este procedimiento no necesariamente satisface las propiedades de optimalidad, pero los resultados indican al menos que en el ámbito de aproximaciones estocásticas se obtienen muy buenos resultados.

Para poder garantizar la eficiencia del estimador, se debe establecer una convergencia en algún sentido de probabilidad del estimador, dado el parámetro de interés,

este hecho tiene sus bases en el siguiente enunciado.

§1.1. Teorema de Convergencia

Supongamos en adelante que $H(y|x)$ es, para todo x , una función de distribución en y , y existe una constante positiva C tal que,

$$Pr[|Y(x)| \leq C] = \int_{-C}^C dH(y|x) = 1, \quad \text{para todo } x. \quad (1.3)$$

Se sigue, en particular que para todo x el valor esperado $M(x)$ definido por (1.2) existe y es finito. Suponiendo, más aún, que existen constantes finitas α, θ tales que

$$M(x) \leq \alpha, \quad \text{para } x < \theta; \quad M(x) \geq \alpha, \quad \text{para } x > \theta. \quad (1.4)$$

Sea $\{a_n\}$ una sucesión fija de constantes positivas tales que

$$0 < \sum_{n=1}^{\infty} a_n^2 = A < \infty. \quad (1.5)$$

Podemos destacar que dicha sucesión representa un papel importante en la aproximación, ya que está determina el tamaño del paso.

Se define una cadena de Markov no estacionaria, $\{x_n\}$, tomando x_1 como una constante arbitraria y definida de la forma

$$x_{n+1} - x_n = a_n(\alpha - y_n), \quad (1.6)$$

donde y_n es una variable aleatoria que representa las observaciones para cada x_n tal que

$$Pr[y_n \leq y | x_n] = H(y|x_n). \quad (1.7)$$

Sea

$$b_n = E(x_n - \theta)^2. \quad (1.8)$$

Encontraremos las condiciones sobre las cuales

$$\lim_{n \rightarrow \infty} b_n = 0, \quad (1.9)$$

sin importar el valor inicial x_1 , que fue definido como un punto arbitrario. Así es bien sabido, (1.9) implica la convergencia en probabilidad de x_n a θ . La expresión dada para b_n , es lo que garantiza el hecho de que el estimador converja a el parámetro y esto es controlado mediante el error cuadrático medio que es fácil de manejar y ofrece una convergencia efectiva del método.

De (1.6) se tiene

$$\begin{aligned} b_{n+1} &= E(x_{n+1} - \theta)^2 = E[E[(x_{n+1} - \theta)^2 | x_n]] \\ &= E \left[\int_{-\infty}^{+\infty} \{(x_n - \theta) - a_n(y - \alpha)\}^2 dH(y|x_n) \right] \\ &= b_n + a_n^2 E \left[\int_{-\infty}^{+\infty} (y - \alpha)^2 dH(y|x_n) \right] - 2a_n E[(x_n - \alpha)(M(x_n) - \alpha)]. \end{aligned}$$

Haciendo

$$d_n = E[(x_n - \alpha)(M(x_n) - \alpha)], \quad (1.10)$$

$$e_n = E \left[\int_{-\infty}^{+\infty} (y - \alpha)^2 dH(y|x_n) \right], \quad (1.11)$$

podemos escribir

$$b_{n+1} - b_n = a_n^2 e_n - 2a_n d_n. \quad (1.12)$$

A partir de (1.4)

$$d_n \geq 0. \quad (1.13)$$

De la ecuación (1.11), tenemos

$$\begin{aligned}
 e_n &= E\left[\int_{-\infty}^{+\infty} (y - \alpha)^2 dH(y|x_n)\right] \\
 &= E\left[\int_{-\infty}^{+\infty} (y^2 - 2\alpha y + \alpha^2) dH(y|x_n)\right] \\
 &\leq E\left[\int_{-\infty}^{+\infty} (C + 2C|\alpha| + |\alpha|^2)\right] \\
 &= [(C + 2C|\alpha| + |\alpha|^2)] = [C + |\alpha|]^2,
 \end{aligned}$$

mientras que de (1.3)

$$0 \leq e_n \leq [C + |\alpha|]^2 < \infty.$$

De lo anterior junto con (1.5) implica que la serie de términos positivos $\sum a_n^2 e_n$ converge, puesto que la serie $\sum a_n^2$ converge a un número finito y además la sucesión e_n está acotada superiormente.

Sumando (1.12) obtenemos

$$\sum_{j=1}^n b_{j+1} - b_j = \sum_{j=1}^n a_j^2 e_j - 2 \sum_{j=1}^n a_j d_j, \quad (1.14)$$

del lado izquierdo de la ecuación anterior, se tiene una suma telescópica que resolviendo da la ecuación

$$b_{n+1} - b_1 = \sum_{j=1}^n a_j^2 e_j - 2 \sum_{j=1}^n a_j d_j. \quad (1.15)$$

Puesto que $b_{n+1} \geq 0$ se tiene que

$$\sum_{j=i}^n a_j d_j \leq \frac{1}{2} \left[b_1 + \sum_{n=1}^{\infty} a_n^2 e_n \right] < \infty. \quad (1.16)$$

Por lo tanto, la serie de términos positivos

$$\sum_{n=1}^{\infty} a_n d_n \quad (1.17)$$

converge. Se sigue de (1.14) que

$$\lim_{n \rightarrow +\infty} b_n = b_1 + \sum_{n=1}^{\infty} a_n^2 e_n - 2 \sum_{n=1}^{\infty} a_n d_n = b \quad (1.18)$$

existe; $b \geq 0$.

Ahora supongamos que existe una sucesión $\{k_n\}$ de constantes no negativas tal que

$$d_n \geq k_n b_n, \quad \sum_{n=1}^{\infty} a_n k_n = \infty. \quad (1.19)$$

De la primera parte de (1.19) y de la convergencia de (1.17) se sigue que

$$\sum_1^{\infty} a_n k_n b_n < \infty. \quad (1.20)$$

De (1.20) y la segunda parte de (1.19) se sigue que para cualquier $\epsilon > 0$ deben existir infinitos valores de n tal que $b_n < \epsilon$. Ya que conocemos que $b = \lim_{n \rightarrow \infty} b_n$ existe, se sigue que $b = 0$.

CAPÍTULO 2

TÉCNICA DE APROXIMACIÓN ESTOCÁSTICA PARA LA ESTIMACIÓN DE PARÁMETROS DE MÁXIMA VEROSIMILITUD

Las aproximaciones estocásticas (SA, sus siglas en ingles) son ampliamente aplicables en técnicas para encontrar raíces de ecuaciones. Estos algoritmos garantizan la convergencia bajo condiciones generalmente más débiles en la forma de la función que es optimizada. La estimación dada no es generalmente óptimal en el sentido estocástico usual (esto es, no necesariamente la varianza es mínima o invariante). En contraste con estas estimaciones subóptimas de aproximaciones estocásticas, se pueden utilizar para generar estimaciones de máxima verosimilitud cuando no es factible o muy difícil la aplicación de algoritmos determinísticos[8].

Supongamos que el estimador de máxima verosimilitud es encontrado por la determinación de las raíces de la ecuación

$$s(\theta) = \frac{\partial L(\theta)}{\partial \theta} = 0, \quad (2.1)$$

donde $L(\cdot)$ es la función de verosimilitud o log- verosimilitud para el parámetro $\theta \in \mathbf{R}^p$. Implementando métodos de descenso rápido o Newton-Raphson, se requiere que $s(\theta)$ y $\partial s / \partial \theta^T$ (Newton- Raphson) o $E(\partial s / \partial \theta^T)$ (scoring) sea evaluado en diversos valores de θ , el algoritmo procede a converger cerca de la raíz de (2.1), digamos θ^* . Un procedimiento que se emplea solo para estimar θ , que requiere que $L(\cdot)$ sea evaluado, es algoritmo de aproximaciones estocásticas de Kiefer and Wolfowitz (KWSA), el cual tiene su forma recursiva dada por

$$z_{n+1} = z_n + a_n \frac{(y_{2n} - y_{2n-1})}{c_n},$$

donde y_{2n-1} y y_{2n} son variables aleatorias independientes con distribuciones respectivamente $H(y|z_n - c_n)$ y $H(y|z_n + c_n)$, y sean $\{a_n\}$ y $\{c_n\}$ sucesiones finitas de números positivos. Lo complicado en la puesta en práctica de KWSA es la aproximación de las diferencias finitas de la derivada, $s(\cdot)$, evaluado en cada valor de θ que se requiere para las iteraciones. Cada aproximación requiere $2p$ evaluaciones de $L(\cdot)$.

El objetivo es establecer como las aproximaciones estocásticas pueden ser utilizadas para formar la estimación de máxima verosimilitud (MLE) con una aproximación de la derivada con excepción de la aproximación estándar de las diferencias finitas usada en KWSA. En particular, esta aproximación alternativa de la derivada requiere sólo dos evaluaciones de $L(\cdot)$, en contraparte a las $2p$ evaluaciones que requiere KWSA. Evaluaremos la compensación entre el cálculo para la iteración asociada con la aproximación alternativa de la derivada y el (esperado) incremento del número de iteraciones a converger (relativo a KWSA), la experiencia numérica indica que esta compensación es favorable cuando p es moderadamente grande.

§2.1. Algoritmo de Aproximaciones Estocásticas y Condiciones de Regularidad Asociada

Las aproximaciones estocásticas son una técnica iterativa para encontrar raíces. Puesto que es interesante encontrar la raíz θ^* de $s(\theta) = 0$. Así lo que se busca es establecer la solución de (2.1) cuando la función de log-verosimilitud no es conocida. Por supuesto, en la estimación de ML, $s(\cdot)$ es conocida para cualquier θ fijo. El objetivo aquí es precisamente reemplazar los cálculos de $s(\cdot)$ (los cuales pueden ser computacionalmente más pesados) con una muy simple aproximación, calculando el estimador. Relacionado al descenso rápido o scoring, esto bajara la carga de cálculos por iteración significativamente, no obstante en el costo probable de un número creciente de iteraciones.

Sea $\hat{\theta}_k$ el estimador para θ^* en la k -ésima iteración, donde $\hat{\theta}_k \in \Lambda(k) \subseteq \mathbf{R}^p$, el algoritmo estándar de SA tiene la forma:

$$\hat{\theta}_k = \hat{\theta}_{k-1} - a_{k-1} \hat{s}(\hat{\theta}_{k-1}), \quad (2.2)$$

donde $\{a_k\}$ es una sucesión creciente que satisface las siguientes condiciones:

$$\lim_{k \rightarrow \infty} a_k = 0, \quad \sum_{k=0}^{\infty} a_k = \infty, \quad \sum_{k=0}^{\infty} a_k^2 < \infty.$$

Observemos que existe una relación estrecha de (2.2) con el método de descenso rápido, la diferencia entre estos métodos radica en que el descenso rápido usa $s(\cdot)$ y en este caso es sustituido por el estimador, $\hat{s}(\cdot)$, que en la k -ésima iteración del método de aproximaciones estocásticas del algoritmo se define como:

$$\hat{s}(\hat{\theta}_k) = \begin{bmatrix} \frac{L[\hat{\theta}_k + \Delta_k] - L[\hat{\theta}_k - \Delta_k]}{2\Delta_{1k}} \\ \vdots \\ \frac{L[\hat{\theta}_k + \Delta_k] - L[\hat{\theta}_k - \Delta_k]}{2\Delta_{pk}} \end{bmatrix}, \quad (2.3)$$

donde $\{\Delta_{1k}, \dots, \Delta_{pk}\}$ es un vector de p variables aleatorias mutuamente independientes.

Veremos aquí el funcionamiento de (2.2) con $\hat{s}(\cdot)$ como se definió antes, y contrastar este funcionamiento con los algoritmos de descenso rápido debido a la estrecha relación que tienen.

La condición principal sobre $\hat{s}(\cdot)$ que son generalmente impuestas por el algoritmo SA en (2.2) son, en términos de el error $e(\hat{\theta}_k) = \hat{s}(\hat{\theta}_k) - s(\hat{\theta}_k)$,

$$E[e(\hat{\theta}_k) | \hat{\theta}_k] = 0 \quad \forall k, \quad (2.4)$$

$$E[\|e(\hat{\theta}_k)\|_2^2 | \hat{\theta}_k] \leq c < \infty \quad \forall k, \quad (2.5)$$

$$\{e(\hat{\theta}_k) | \hat{\theta}_k\}_{k=0}^{\infty} \quad \text{mutuamente independiente,} \quad (2.6)$$

donde $\|\cdot\|_2$ denota la norma L^2 (Euclidea). Bajo este y las condiciones mencionadas antes de a_k , junto con cierta condición de regularidad (ver, [4] o Kushner y Clark [3]), $\hat{\theta}(k) \rightarrow \theta^*$ cuando $k \rightarrow \infty$. Las esperanzas y probabilidades planteadas están condicionadas sobre los datos que aparecen en la función de verosimilitud.

§2.2. Estimador, $\hat{s}(\cdot)$

Sea $\Delta_k \in \Omega(k) \subseteq \mathbf{R}^p$, y $\{\Delta_{1k}, \dots, \Delta_{pk}\}$ un vector de p variables aleatorias mutuamente independientes. Además sea $\{\Delta_k\}_{k=1}^{\infty}$ una sucesión mutuamente independiente.

Consideremos las condiciones (2.5) y (2.6). La ecuación (2.6) claramente se satisface debido a que Δ_k es mutuamente independiente. Ahora consideraremos la condición de segundo momento, (2.5). Por ahora, sea k fijo (y así el argumento k es suprimido). Sea $\Gamma \subseteq \mathbf{R}^p$ y supongamos que $\hat{\theta} \pm \Delta \in \Gamma$ y $s(\theta)$ es continuo y acotado en Γ , por el Teorema de Valor Medio se tiene que:

$$L(\hat{\theta} \pm \Delta) = L(\hat{\theta}) + s(\bar{\theta}^{\pm})^T (\pm \Delta)$$

donde $\bar{\theta}^{\pm}$ denota un punto en el segmento de recta entre $\hat{\theta}$ y $\hat{\theta} \pm \Delta$ (apropiado). Así la i -ésima componente de \hat{s} satisface

$$\hat{s}_i(\hat{\theta}) = \frac{[s(\bar{\theta}^+) - s(\bar{\theta}^-)]^T \Delta}{\Delta_i},$$

con $|\hat{s}_i(\hat{\theta})| \leq c' \sum_{j=1}^p |\Delta_j / \Delta_i|$ donde $c' = 2 \max_j \sup_{\theta \in \Gamma} \|s_j(\theta)\|$, la cual está bien definida (acotado) debido a que $s(\theta)$ está acotado en Γ . Así por la Desigualdad de Minkowski,

$$E(\|\hat{s}(\hat{\theta})\|_2^2 | \hat{\theta}) \leq (c')^2 \sum_{i=1}^p \left\{ \sum_{j=1}^p \left[E\left(\frac{\Delta_j}{\Delta_i}\right)^2 \right]^{1/2} \right\}^2, \quad (2.7)$$

la cual, por supuesto, es acotada cuando $E(\Delta_j/\Delta_i)^2 \leq c' < \infty$ para todo i, j (cuando Δ_i tiene una Distribución Bernoulli con resultados distintos de cero). Ahora, (2.5) se satisface para algún k si la cota en (2.7) es finita para $\hat{\theta} = \hat{\theta}_k$, recuerde que $s(\hat{\theta}_k)$ es considerado una constante relativa a la medida $P(\cdot|\hat{\theta}_k)$. Así (2.5) se satisface si $s(\theta)$ es continua y acotada sobre $\bigcup_{k=0}^{\infty} \Gamma(k)$ y si $E(\Delta_{jk}/\Delta_{ik})^2 \leq c' < \infty \forall i, j = 1, 2, \dots, p; k = 1, 2, \dots, \infty$.

§2.3. Sesgo del estimador $\hat{s}(\cdot)$

El estimador $\hat{s}(\cdot)$ dado por la ecuación (2.3), bajo las consideraciones con que fue definido veremos mediante el teorema que se dará a continuación que es un estimador insesgado de $s(\cdot)$, con un término de error de $O(\delta^2)$, donde δ es alguna constante positiva que puede ser arbitrariamente pequeña sujeta a la exactitud en las limitaciones del computador. En lo que sigue estableceremos los Teoremas que permiten definir esa característica del estimador.

El teorema que será presentado establece una cota para el sesgo del estimador $E[e(\hat{\theta})|\hat{\theta}] = E[\hat{s}(\hat{\theta}) - s(\hat{\theta})|\hat{\theta}]$ asumiendo que $L(\theta)$ es de clase C^∞ en la región sobre $\hat{\theta}$.

§2.3.1. Teorema (Cota del sesgo del estimador)

Supongamos que Δ_j está simétricamente distribuida sobre 0, $|\Delta_j| \leq \delta$ a.s., y $|\Delta_j|^{-1} \leq M(\delta) = O(\delta^{-2})$ a.s. ($\delta \rightarrow 0$) $\forall j = 1, 2, \dots, p$. Además, supongamos que $L^{(m)}(\hat{\theta} \pm \Delta)$ existe $\forall m = 1, 2, \dots$, y

$$|L_{i_1, i_2, \dots, i_m}^{(m)}(\hat{\theta} \pm \Delta)| \leq \beta(m) \quad (2.8)$$

sobre Ω para cualquier $\hat{\theta} \in \Lambda$ donde

$$\beta(m) = O(m! / (\delta p + \epsilon)^m) \quad (m \rightarrow \infty) \quad (2.9)$$

para cualquier $\epsilon > 0$. Entonces $\forall j = 1, 2, \dots, p$,

$$\begin{aligned}
 |E[e_j(\hat{\theta})|\hat{\theta}]| &\leq b_j(\delta) \\
 &\equiv \delta^{-1} \sum_{i=1}^{\infty} [(p\delta)^{2i+1} - ((p-1)\delta)^{2i+1}] \frac{\beta(2i+1)}{(2i+1)!} \quad (2.10)
 \end{aligned}$$

$$= O(\delta^2) \quad (\delta \rightarrow 0). \quad (2.11)$$

Demostración:

(2.8) y (2.9) implica que el resto de una Expansión de Taylor de orden m de $L(\hat{\theta} \pm \Delta)$ sobre $\hat{\theta}$ satisface

$$|L(\hat{\theta} \pm \Delta) - T^{(m)}(\hat{\theta} \pm \Delta)| \leq \frac{(p\delta)^{m+1}}{(m+1)!} \beta(m+1) \rightarrow 0$$

cuando $m \rightarrow \infty$, lo cual implica que

$$\hat{s}_l(\hat{\theta}) = \frac{\sum_{i=0}^{\infty} L^{(2i+1)}(\hat{\theta}) \Delta^{[2i+1]} / (2i+1)!}{\Delta_l}.$$

Ahora, ya que $E(\Delta_j / \Delta_i) = 0 \forall j \neq l$ entonces,

$$E[\hat{s}_l(\hat{\theta})|\hat{\theta}] = s_l(\hat{\theta}) + E[e_l(\hat{\theta})|\hat{\theta}],$$

donde,

$$E[e_l(\hat{\theta})|\hat{\theta}] = E\left[\frac{\sum_{i=1}^{\infty} L^{(2i+1)}(\hat{\theta}) \Delta^{[2i+1]} / (2i+1)!}{\Delta_l} \mid \hat{\theta}\right]. \quad (2.12)$$

Note que para cualquier $m \in \{3, 5, 7, \dots\}$, $(p-1)^m$ (fuera de un total de términos de p^m) los términos que no contienen a Δ_l . Así, usando (2.8) y el hecho que los Δ_i están simétricamente distribuidas, conocemos que para $1 \leq i_1, i_2, \dots, i_m \leq p$,

$$E\left[\frac{L^{(m)}_{i_1, i_2, \dots, i_m} \Delta_{i_1} \Delta_{i_2} \cdots \Delta_{i_m}}{\Delta_l} \mid \hat{\theta}\right]$$

$$\begin{cases} = 0 & \text{para } (p-1)^m \text{ términos;} \\ \leq \beta(m)\delta^{m-1} & \text{en magnitud para } p^m - (p-1)^m \text{ términos.} \end{cases}$$

(donde el punto de evaluación $\hat{\theta}$ ha sido suprimido en $L^{(m)}$), lo cual implica que

$$|E(L^{(m)} \Delta^{[m]} / \Delta_l | \hat{\theta})| \leq [p^m - (p-1)^m] \beta(m) \delta^{m-1}. \quad (2.13)$$

Ahora definimos

$$D(\delta) = M(\delta) \sum_{i=1}^{\infty} \beta(2i+1) (p\delta)^{2i+1} / (2i+1)!,$$

la cual es finito por (2.9). Entonces de (2.8) y del hecho que $|\Delta_l|^{-1} \leq M(\delta)$, sabemos que

$$\left| \frac{\sum_{i=1}^m L^{(2i+1)} \Delta^{[2i+1]} / (2i+1)!}{\Delta_l} \right| \leq D(\delta) \quad (2.14)$$

$\forall m = 1, 2, \dots$ Además, aplicando la Regla de L'Hopital a $D(\delta)$ produce $D(\delta) = O(1)$ ($\delta \rightarrow 0$), y tal $D(\delta)$ es integrable ($\int_{\Omega} D(\delta) dP_{\Delta} = D(\delta)$) cuando $\delta \rightarrow 0$. Así de (2.14) y por el Teorema de Convergencia Dominante de Lebesgue, la esperanza y suma sobre (2.12) puede ser intercambiado, lo cual por (2.13) produce (2.10).

Ahora, demostremos (2.11), aplicando la regla de L'Hopital a (2.10), donde encontramos que

$$\begin{aligned} b_l(\delta) &\sim [p^3 - (p-1)^3] \beta(3) \delta^2 / 6 \\ &= O(\delta^2), \end{aligned}$$

lo cual termina la prueba puesto que l fue elegido arbitrariamente.

§2.4. Modificación de segundo orden del algoritmo de SPSA

El desarrollo del Método de Aproximaciones Estocásticas y Perturbaciones Sucesivas (SPSA) nos permite encontrar muchas aplicaciones en la estimación de parámetros. Existen dos algoritmos básicos SPSA que están basados en el concepto de perturbaciones sucesivas (SP) que solo usan mediciones (ruido) sobre la función de pérdida. SPSA de primer orden (1SPSA) esta relacionado al método de aproximaciones estocásticas de Kiefer- Wolfowitz (K-W), mientras que SPSA de segundo orden (2SPSA) es una analogía del algoritmo determinístico de Newton- Raphson. El método

de 1SPSA es superior a otros métodos de primer orden de Aproximaciones Estocásticas (SA), debido a la eficiencia de este estimador para el cálculo del gradiente de la función de objetivo.

Está demostrado que la implementación "estándar" de 2SPSA alcanza un error asintótico óptimo, con el error cuadrático medio siendo este no más que dos veces el error óptimo de una sucesión creciente dependiendo de la tercera derivada de la función objetivo [11].

Una característica importante entre estos dos métodos puede ser vista desde la perspectiva de la inclusión y condicionamiento de la matriz Hessiana de la función.

Los algoritmos de SPSA tienen como forma base la recurrencia general para la estimación de $(\hat{\theta}_k)$ de una solución (θ^*) de dimensión p , en las siguientes forma:

1SPSA:

$$\hat{\theta}_{k+1} = \hat{\theta}_k - a_k \hat{g}_k(\hat{\theta}_k), \quad k = 0, 1, 2, \dots \quad (2.15)$$

2SPSA:

$$\hat{\theta}_{k+1} = \hat{\theta}_k - \bar{a}_k \bar{H}_k^{-1} \hat{g}_k(\hat{\theta}_k), \quad \bar{H}_k = f_k(\bar{H}_k) \quad (2.16)$$

$$\bar{H}_k = \frac{k}{k+1} \bar{H}_{k-1} + \frac{1}{k+1} \hat{H}_k, \quad k = 0, 1, 2, \dots \quad (2.17)$$

donde a_k y \bar{a}_k son sucesiones escalares crecientes que ciertamente satisfacen la condición de SA, \hat{g}_k es el estimador de SP del gradiente de la función de pérdida que depende de la sucesión creciente c_k (representando una diferencia del intervalo de la perturbación), \hat{H}_k es el estimador de SP de la matriz Hessiana, y f_k es la aplicación donde generalmente \bar{H}_k es una matriz $p \times p$ definida positiva o no. Sea Δ_k definido como un vector aleatorio con media cero y de dimensión p , donde cada componente es una variable aleatoria independiente.

El i -ésimo elemento del gradiente de la función de pérdida esta dado por

$$(\hat{g}_k)_i = (2c_k\Delta_{ki})^{-1}[y(\hat{\theta}_k + c_k\Delta_k) - y(\hat{\theta}_k - c_k\Delta_k)], \quad i = 1, 2, \dots, p \quad (2.18)$$

donde Δ_{ki} es la i -ésima componente de Δ_k , el vector $y(\theta)$ son las mediciones de la función de perdida : $y(\theta) = L(\theta) + ruido$, donde θ es el parámetro que tiene como valor real θ^* . Notemos que la forma de 2SPSA es un caso especial de la adaptación del método de SP de Spall [8].

La verdadera matriz hessiana de la función de perdida $H(\theta)$ tiene definido su ij -ésimo elemento como $H_{ij} = \partial^2 L / \partial \theta_i \partial \theta_j$ su valor es la solución de $H(\theta^*)$ denotado por H^* . El ij -ésimo elemento de estimación por iteración del estimador de H esta dado por:

$$(\hat{H}_k)_{ij} = (4c_k\tilde{c}_k)^{-1} + [(\Delta_{ki}\tilde{\Delta}_{kj})^{-1} + (\Delta_{kj}\tilde{\Delta}_{ki})^{-1}][y(\hat{\theta}_k + c_k\Delta_k + \tilde{c}_k\tilde{\Delta}_k) - y(\hat{\theta}_k + c_k\Delta_k) - y(\hat{\theta}_k - c_k\Delta_k + \tilde{c}_k\tilde{\Delta}_k) + y(\hat{\theta}_k - c_k\Delta_k)], \quad i, j = 1, 2, \dots, p \quad (2.19)$$

donde la sucesión \tilde{c}_k satisface condiciones similares a las de c_k y con $\tilde{\Delta}_k = (\tilde{\Delta}_{k1}, \tilde{\Delta}_{k2}, \dots, \tilde{\Delta}_{kp})^T$ es generado de forma similar a el estadístico Δ_k . Note que el estimador del Hessiano es simétrico esto es conveniente en la aplicación de optimización y es un requerimiento crucial para la nueva aplicación f_k propuesta a continuación.

§2.4.1. Aplicación f_k para 2SPSA

Un aspecto relevante en la implementación de 2SPSA es la definición de la aplicación f_k , de \bar{H}_k a \tilde{H}_k sin embargo a menudo en la práctica se puede encontrar que estas matrices pueden ser definidas positiva o no positivas. Para la existencia de mínimo(s) de una función es necesario garantizar que la matriz hessiana sea definida positiva, por esto es presentado un procedimiento que elimina la definición no positiva mientras que preserva las características espectrales dominantes de \tilde{H}_k . Esta aproximación es motivada en ejemplos finitos, como indicaremos en la siguiente discusión. Primero calcularemos los valores propios de \tilde{H}_k y los ordenaremos en orden decreciente:

$$\Lambda_k \equiv \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_{q-1}, \lambda_q, \lambda_{q+1}, \dots, \lambda_p] \quad (2.20)$$

donde $\lambda_q > 0$ y $\lambda_{q+1} \leq 0$. Por simplicidad omitiremos el subíndice k para los valores propios λ_i que están en función de k . Supongamos que los valores propios negativos no serán considerados en la solución, ya que estos son causal de cualquier error en \bar{H}_k debido a que en la iteración podría no ser alcanzada la vecindad de θ^* donde la función de pérdida es localmente cuadrática. Por lo tanto, reemplazamos los valores propios pequeños por una serie decreciente de valores propios positivos:

$$\hat{\lambda}_q = \varepsilon \hat{\lambda}_{q-1}, \hat{\lambda}_{q+1} = \varepsilon \hat{\lambda}_q, \dots, \hat{\lambda}_p = \varepsilon \hat{\lambda}_{p+1}, \quad (2.21)$$

donde el parámetro de ajuste $0 < \varepsilon < 1$ puede ser especificado basado en la existencia de valores propios positivos

$$\varepsilon = (\lambda_{q-1}/\lambda_1)^{q-2}. \quad (2.22)$$

La propuesta de redefinir los valores propios (λ_q) pequeños evita la posibilidad de tener valores cero y así tomar aplicaciones que tengan una matriz hessiana no singular. Sea $\hat{\Lambda}_k$ es la diagonal de la matriz Λ_k , con valores propios $\lambda_q, \dots, \lambda_p$ reemplazados por $\hat{\lambda}_q, \dots, \hat{\lambda}_p$ definidos en concordancia con (2.17).

Las ecuaciones (2.17) y (2.18) indican que la característica espectral de la existencia de valores propios positivos como las mediciones para el radio máximo o mínimo de estos valores propios. De esta forma \bar{H}_k es simétrica, es ortogonal similar a la matriz diagonal inicial de los valores propios reales [5],

$$\bar{H}_k = P_k \Lambda_k P_k^T \quad (2.23)$$

donde P_k es una matriz ortogonal que se forma de todos los vectores propios de \bar{H}_k , la cual usualmente deriva de los valores propios. Ahora, la aplicación f_k puede ser expresada como

$$f_k(\bar{H}_k) = P_k \hat{\Lambda}_k P_k^T. \quad (2.24)$$

Puesto que $\bar{\bar{H}}_k^{-1}$ es usado en la recursión (2.13) de 2SPSA aplicando (2.21) con los vectores propios disponibles de \bar{H}_k también lleva a una inversión fácil para el estimador del Hessiano:

$$\bar{\bar{H}}_k^{-1} = P_k \hat{\Lambda}_k^{-1} P_k^T. \quad (2.25)$$

2SPSA está basado sobre la aplicación (2.21) realiza el procedimiento de eliminación de definiciones no positivas de \bar{H}_k . Se observa que los parámetros dominantes para la aplicación (ϵ y λ_{q-1}) son internamente determinados por \bar{H}_k en cada iteración. Esto es diferente de algunas otras formas de f_k donde es necesario usar un coeficiente específico.

Notemos que en 2SPSA la recursión (2.13) efectivamente involucra cálculos de la inversa de la matriz $\bar{\bar{H}}_k^{-1}$ (aunque, en la implementación, la inversión explícita podría ser evitado usando métodos estándar del álgebra lineal). La aplicación f_k definida por (2.21) garantiza que $\bar{\bar{H}}_k$ es una matriz no singular. Dicha aplicación produce el reemplazo de posibles λ_q ceros por un $\hat{\lambda}_q$ mejor comportado también elimina la posibilidad de una matriz no singular. Sin embargo, los elementos de \bar{H}_k resulta desde la aproximación SP y mediciones imperfectas de la función de perdida son sujeto a errores. Esos errores estarán directamente afectados al cálculo de la matriz inversa.

CAPÍTULO 3

MODELO MACROECONÓMICO

El desarrollo fundamental a la base de la macroeconomía moderna es el llamado Modelo de Crecimiento Neoclásico; en estos modelos las decisiones de los agentes son modelados de manera explícita en un contexto intertemporal.

En este trabajo usaremos un modelo de crecimiento Neoclásico propuesto por Cass y Koopman [10]. Es importante destacar que el modelo básico de crecimiento deja de lado importantes aspectos relacionados con el modelado de una pequeña economía abierta, y nos basaremos en este ya que proporciona un punto de partida a la hora de contrastar e interpretar los cambios que, en la estimación de los parámetros, podrían ocasionar una modificación de la especificación del modelo.

Este modelo esta dado en términos de un agente representativo en la economía cuyas preferencias sobre el consumo C_t y el ocio $(1 - L_t)$ están representados por la función de utilidad:

$$\max_{\{C,L\}} E_0 \left\{ \sum_{t=0}^{\infty} \beta^t \left[\frac{[C_t^\theta (1 - L_t)^{1-\theta}]^{1-\tau}}{1 - \tau} \right] \right\} \quad (3.1)$$

sujeto a las restricciones:

$$C_t + I_t = Z_t K_t^\alpha L_t^{1-\alpha} \quad (3.2)$$

$$K_{t+1} = I_t + (1 - \delta)K_t \quad (3.3)$$

$$Z_t = Z_{ss}^{1-\rho} Z_{t-1}^\rho e^{\xi_t}, \quad \xi_t \sim N(0, \sigma^2) \quad (3.4)$$

donde:

$\beta \in (0, 1)$ es el factor de descuento,

$\theta \geq 0$ es el parámetro que controla la oferta de trabajo,
 $\tau \geq 1$ es la elasticidad de sustitución intertemporal,
 $\alpha \in [0, 1]$ es el parámetro de participación de capital,
 $\delta \geq 0$ es la tasa de depreciación,
 $\rho \in [0, 1]$ es la persistencia del shock tecnológico,
 L_t es la cantidad de trabajo agregado,
 I_t es la inversión,
 Z_t es el nivel de tecnología,
 K_t es el capital agregado.

Definido el modelo tenemos que las variables asociadas al mismo son: C_t, L_t, I_t, K_t, Z_t y los parámetros presentes están dados por: $\gamma = \{\theta, \rho, \alpha, \tau, \delta, \beta, \sigma_\varepsilon\}$

Dado este modelo por su sola naturaleza representan un objetivo de investigación, en el cual se requiere una solución del modelo bajo un cierto criterio de manera teórica, esto puede venir dado al desarrollar un procedimiento descrito por Uhlig para modelos dinámicos estocásticos no lineales.

Este procedimiento tiene su esencia en la log- linealización del modelo original; el procedimiento general que este método sigue para resolver y analizar este tipo de modelos lo seguiremos paso a paso.

A grandes rasgos el método de Uhlig sigue el siguiente esquema; inicialmente se debe plantear el modelo como un problema de optimización, bien definido, es decir, la función objetivo y las restricciones. Después el modelo debe ser expresado como una función Lagrangeana; la ecuación Lagrangeana es un método que reduce el problema restringido de n variables en uno sin restricciones con $n + m$ variables, donde m indica el número de restricciones. El multiplicador de Lagrange que es introducido por la función Lagrangeana, por la naturaleza del problema es considerado uno para cada periodo de tiempo t .

Esta estructura Lagrangeana nos dará un paso importante para la caracteriza-

ción, puesto que se debe tener en consideración que el problema planteado es no lineal y dinámico, lo cual como se ha mencionado nos proporciona una gran dificultad a la hora de establecer una solución y la representación lagrangeana nos ayuda a dar un enfoque lineal que proporciona una teoría mucho más amplia para el estudio.

§3.1. Condiciones de Optimalidad

$$\begin{aligned} \text{máx } f(C_t, L_t, K_{t+1}, \lambda_t) = E_0 \left\{ \sum_{t=0}^{\infty} \beta^t \left[\frac{[C_t^\theta (1 - L_t)^{1-\theta}]^{1-\tau}}{1 - \tau} - \lambda_t (C_t - Z_t K_t^\alpha L_t^{1-\alpha} \right. \right. \\ \left. \left. + K_{t+1} - (1 - \delta)K_t) \right] \right\}. \end{aligned} \quad (3.5)$$

De la ecuación en forma Lagrangeana del modelo presentada antes, podemos establecer las condiciones de optimalidad de primer orden:

$$\frac{\partial f}{\partial C_t} = 0 : \beta^t \theta C_t^{\theta-1} (1 - L_t)^{1-\theta} [C_t^\theta (1 - L_t)^{1-\theta}]^{-\tau} - \beta^t \lambda_t = 0 \quad (3.6)$$

$$\Rightarrow \lambda_t = \theta C_t^{(1-\tau)\theta-1} (1 - L_t)^{(1-\theta)(1-\tau)}. \quad (3.7)$$

$$\begin{aligned} \frac{\partial f}{\partial L_t} = 0 : -\beta^t (1 - \theta) [C_t^\theta (1 - L_t)^{1-\theta}]^{-\tau} (1 - L_t)^{-\theta} + \beta^t \lambda_t (1 - \alpha) L_t^{-\alpha} Z_t K_t^\alpha = 0 \\ = -\beta^t C_t^{(1-\tau)\theta} (1 - L_t)^{-\tau - (1-\tau)\theta} (1 - \theta) + \beta^t \lambda_t (1 - \alpha) Z_t K_t^\alpha L_t^{-\alpha} = 0. \end{aligned} \quad (3.8)$$

$$\frac{\partial f}{\partial K_{t+1}} = 0 : -\lambda_t \beta^t + E_t [\lambda_{t+1} \beta^{t+1} (Z_{t+1} \alpha K_{t+1}^{\alpha-1} L_{t+1}^{1-\alpha} + 1 - \delta)] = 0 \quad (3.9)$$

$$\Rightarrow \lambda_t = \beta E_t [\lambda_{t+1} (Z_{t+1} \alpha K_{t+1}^{\alpha-1} L_{t+1}^{1-\alpha} + 1 - \delta)] = 0. \quad (3.10)$$

$$\frac{\partial f}{\partial \lambda} = 0 : \beta^t [C_t - Z_t K_t^\alpha L_t^{1-\alpha} + K_{t+1} - (1 - \delta)K_t] = 0. \quad (3.11)$$

$$\log(Z_t) = (1 - \rho) \log(Z_{ss}) + \rho \log(Z_{t-1}) + \xi_t, \quad \xi_t \sim N(0, \sigma^2), |\rho| < 1. \quad (3.12)$$

Estas ecuaciones de optimalidad pueden ser representadas mediante el siguiente sistema reducido:

1. Condición de Euler, se obtiene reemplazando la ecuación (3.7) en la (3.10):

$$\frac{[C_t^\theta(1-L_t)^{1-\theta}]^{1-\tau}}{C_t} = \beta E_t \left\{ \frac{[C_{t+1}^\theta(1-L_{t+1})^{1-\theta}]^{1-\tau}}{C_{t+1}} (1 + \alpha Z_{t+1} K_{t+1}^{\alpha-1} L_{t+1}^{1-\alpha} - \delta) \right\}. \quad (3.13)$$

2. Condición Intratemporal (sustituyendo (3.7) en (3.8)):

$$\frac{[C_t^\theta(1-L_t)^{1-\theta}]^{1-\tau}}{1-L_t} = \frac{\theta}{(1-\theta)} (1-\alpha) \frac{[C_t^\theta(1-L_t)^{1-\theta}]^{1-\tau}}{C_t} Z_t K_t^\alpha L_t^{-\alpha}. \quad (3.14)$$

3. Restricción Agregada de Recurso (por (3.9)):

$$C_t + K_{t+1} = Z_t K_t^\alpha L_t^{1-\alpha} + (1-\delta)K_t. \quad (3.15)$$

4. Shock Tecnológico (por (3.12)) asumiendo que $Z_{ss} = 1$:

$$\log(Z_t) = \rho \log(Z_{t-1}) + \xi_t, \quad \xi_t \sim N(0, \sigma^2), |\rho| < 1. \quad (3.16)$$

§3.1.1. Estado Estacionario

El estado estacionario radica en suponer que las variables son invariantes en el tiempo, esto no es más que eliminar la dependencia con el tiempo omitiendo el subíndice t . En este caso se usan subíndices ss para denotar dicho estado.

Como todas las variables son constantes y además se requiere que el estado futuro sea igual al actual, esto es

$$\{C_t = C_{t+1}, K_t = K_{t+1}, L_t = L_{t+1}, Z_t = Z_{t+1}\}.$$

El estado estable o estacionario del modelo básico de crecimiento puede ser descrito mediante cada una de las ecuaciones planteadas a continuación:

1. Estado estacionario a partir de la Condición de Euler:

$$1 = \beta(1 + \alpha K_{ss}^{\alpha-1} L_{ss}^{1-\alpha} - \delta). \quad (3.17)$$

2. Estado estacionario derivado a partir de la Condición Intratemporal:

$$\frac{1}{1 - L_{SS}} = \frac{\theta(1 - \alpha) K_{SS}^{\alpha} L_{SS}^{-\alpha}}{(1 - \theta) C_{SS}}. \quad (3.18)$$

3. Estado estacionario derivado a partir de la Restricción Agregada de Recursos:

$$C_{SS} + K_{SS} = K_{SS}^{\alpha} L_{SS}^{1-\alpha} + (1 - \delta)K_{SS}. \quad (3.19)$$

4. Estado estacionario a partir del Shock Tecnológico:

$$Z_{SS} = 0. \quad (3.20)$$

Determinado el estado estacionario de las variables procedemos a verificar de que parámetro dependen las variables.

Resolviendo el sistema de ecuaciones, se tiene:

$$L_{SS} = \frac{1}{\Omega + 1} \quad (3.21)$$

$$K_{SS} = \Theta^{\frac{1}{\alpha-1}} L_{SS} \quad (3.22)$$

$$C_{SS} = (\Theta - \delta)K_{SS} \quad (3.23)$$

$$Y_{SS} = K_{SS}^{\alpha} L_{SS}^{1-\alpha}, \quad (3.24)$$

donde:

$$\Theta = \frac{Y_{SS}}{K_{SS}} = \frac{1}{\alpha} \left(\frac{1}{\beta} - 1 + \delta \right), \quad (3.25)$$

$$\Omega = \frac{(1 - \theta)}{(1 - \alpha)\theta} \left(1 - \frac{\delta}{\Theta} \right) \quad (3.26)$$

§3.1.2. Log- Linealización del Modelo

El modelo estándar de aproximación lineal consiste en aproximar alrededor del estado de equilibrio, donde esto es válido si y sólo si el modelo no está muy lejos de su estado estacionario, para obtener resultados al modelo original.

Esta técnica esta basada en usar la aproximación de Taylor alrededor del estado estacionario y se reemplazan todas las ecuaciones por aproximaciones que son funciones lineales en las log- linealizaciones de las variables.

La estrategia es usar una aproximación de Taylor de primer orden alrededor del estado estable y reemplazar las ecuaciones con aproximaciones, las cuales son lineales en las log-desviaciones de las variables.

Sea X_t una variable estrictamente positiva, X_{ss} es el estado estable y

$$x_t = \log(X_t) - \log(X_{ss}) \quad (3.27)$$

es la desviación logarítmica.

Notemos primero que, para X_{ss} pequeño, $\log(1 + X_{ss}) \simeq X_{ss}$, así

$$X_t = \log(X_t) - \log(X_{ss}).$$

Este método puede ser expresado de la siguiente forma. Supongamos que se tiene una ecuación de la siguiente forma:

$$f(X_t, Y_t) = g(Z_t), \quad (3.28)$$

donde X_t, Y_t y Z_t son variables estrictamente positivas.

Esta ecuación es claramente válida para el estado- estable

$$f(X_{ss}, Y_{ss}) = g(Z_{ss}). \quad (3.29)$$

Encontremos la log- linealización de la ecuación (3.28), reescribimos las variables usando la identidad $X_t = e^{\log(x_t)}$ y entonces tomamos logaritmos en ambos lados:

$$\log(f(e^{\log(X_t)}, e^{\log(Y_t)})) = \log(g(e^{\log(Z_t)})). \quad (3.30)$$

Ahora tomemos la aproximación de Taylor de primer orden alrededor del estado-estable $(\log(X_{ss}), \log(Y_{ss}), \log(Z_{ss}))$. Después de algunos cálculos, podemos escribir el lado izquierdo como

$$\begin{aligned} \log(f(X_{ss}, Y_{ss})) + \frac{1}{f(X_{ss}, Y_{ss})} [f_1(X_{ss}, Y_{ss}) X_{ss} (\log(X_t) - \log(X_{ss})) \\ + f_2(X_{ss}, Y_{ss}) Y_{ss} (\log(Y_t) - \log(Y_{ss}))]. \end{aligned} \quad (3.31)$$

Simultáneamente, el lado derecho puede escribirse como

$$\log(g(Z_t)) + \frac{1}{g(Z_{ss})} [g'(Z) Z_{ss} (\log(Z_t) - \log(Z_{ss}))]. \quad (3.32)$$

De las ecuaciones (3.31) y (3.32), y usando (3.27) y (3.29), produce la siguiente ecuación de log-linealización

$$[f_1(X_{ss}, Y_{ss}) X_{ss} x_t + f_2(X_{ss}, Y_{ss})] \simeq [g'(Z_{ss}) Z_{ss} z_t]. \quad (3.33)$$

Notemos que esta última es la ecuación lineal en las desviaciones.

Ahora bajo un procedimiento similar al que se planteo anteriormente representaremos cada una de las ecuaciones del modelo en su forma log-linealizada.

1. Log-linealización de la Condición de Euler:

$$\begin{aligned} (\theta(1 - \tau) - 1) \hat{C}_t - (1 - \tau)(1 - \theta) \frac{L_{ss}}{1 - L_{ss}} \hat{L}_t = E_t \{ (\theta(1 - \tau) - 1) \hat{C}_{t+1} \\ - (1 - \tau)(1 - \theta) \frac{L_{ss}}{1 - L_{ss}} \hat{L}_{t+1} \\ + \alpha \beta K_{ss}^{\alpha-1} L_{ss}^{1-\alpha} (\hat{Z}_{t+1} + (\alpha - 1) \hat{K}_{t+1} \\ + (1 - \alpha) \hat{L}_{t+1}) \}. \end{aligned} \quad (3.34)$$

2. Log- linealización de la Condición Intratemporal:

$$\hat{C}_t + \frac{L_{ss}}{1 - L_{ss}} \hat{L}_t = \hat{Z}_t + \alpha \hat{K}_t - \alpha \hat{L}_t. \quad (3.35)$$

3. Log- linealización de la restricción Agregada de Recursos:

$$C_{ss} \hat{C}_t + K_{ss} \hat{K}_{t+1} = K_{ss}^\alpha L_{ss}^{1-\alpha} (\hat{Z}_t + \alpha \hat{K}_t + (1 - \alpha) \hat{L}_t) + (1 - \delta) K_{ss} \hat{K}_t. \quad (3.36)$$

4. Log linealización del Shock Tecnológico:

$$\hat{Z}_t = \rho \hat{Z}_{t-1} + \xi_t. \quad (3.37)$$

§3.1.3. Resolución del modelo

Para poder resolver el modelo es necesario reescribir las condiciones log- linealizadas descritas anteriormente, como:

$$E_t \{ \alpha_1 \hat{C}_{t+1} + \alpha_2 \hat{L}_{t+1} + \alpha_3 \hat{Z}_{t+1} - \alpha_4 \hat{K}_{t+1} + \alpha_4 \hat{L}_{t+1} - \alpha_1 \hat{C}_t + \alpha_2 \hat{L}_t \} = 0 \quad (3.38)$$

$$\hat{C}_t + \alpha_5 \hat{L}_t - \hat{Z}_t - \alpha \hat{K}_t = 0 \quad (3.39)$$

$$C_{ss} \hat{C}_t + K_{ss} \hat{K}_{t+1} - Y_{ss} \hat{Z}_t - Y_{ss} \alpha \hat{K}_t - Y_{ss} (1 - \alpha) \hat{L}_t - (1 - \delta) K_{ss} \hat{K}_t = 0 \quad (3.40)$$

$$\hat{Z}_t = \rho \hat{Z}_{t-1} + \xi_t, \quad (3.41)$$

donde: $\alpha_1 = \theta(1 - \tau)$, $\alpha_2 = -(1 - \tau)(1 - \theta) \frac{L_{ss}}{1 - L_{ss}}$, $\alpha_3 = \alpha \beta K_{ss}^{\alpha-1} L_{ss}^{1-\alpha}$, $\alpha_4 = \alpha_3(1 - \alpha)$, $\alpha_5 = (\frac{L_{ss}}{1 - L_{ss}} + \alpha)$, $Y_{ss} = K_{ss}^\alpha L_{ss}^{1-\alpha}$, \hat{K}_{t+1} es un vector de estado endógeno de dimensión $m \times 1$, \hat{L}_t un vector de otras variables endógenas $n \times 1$ y Z_t un vector de variables exógenas $k \times 1$.

Las relaciones de equilibrio entre estas variables se representan por el siguiente sistema, propuesto por Uhlig:

$$A \hat{K}_{t+1} + B \hat{K}_t + C \hat{L}_t + D \hat{Z}_t = 0 \quad (3.42)$$

$$E_t \{ G \hat{K}_{t+1} + H \hat{K}_t + J \hat{L}_{t+1} + K \hat{L}_t + L \hat{Z}_{t+1} + M \hat{Z}_t \} = 0 \quad (3.43)$$

$$E_t(\hat{Z}_{t+1}) = N \hat{Z}_t, \quad (3.44)$$

donde $\mathbf{A} = K_{SS}$, $\mathbf{B} = \alpha(C_{SS} - Y_{SS}) - (1 - \delta)K_{SS}$, $\mathbf{C} = Y_{SS}(\alpha - 1) - \alpha_5 C_{SS}$, $\mathbf{D} = C_{SS} - Y_{SS}$,
 $\mathbf{G} = (\alpha_1 \alpha - \alpha_4)$, $\mathbf{H} = -\alpha_1 \alpha$, $\mathbf{J} = \alpha_4 - \alpha_1 \alpha_5 + \alpha_2$, $\mathbf{K} = \alpha_2 + \alpha_1 \alpha_5$, $\mathbf{L} = \alpha_3 + \alpha_1$,
 $\mathbf{M} = -\alpha_1$, $\mathbf{N} = \rho$.

Se considera que \mathbf{C} es de dimensión $l \times n$, $l \geq n$ y de rango n , \mathbf{F} es de tamaño $(m + n - 1) \times n$ y que tiene N autovalores estables.

La solución propuesta para el sistema de ecuaciones dado por (3.42), (3.43) y (3.44) es la ley de movimiento de equilibrio recursivo, usada en Uhlig:

$$\hat{K}_{t+1} = \mathbf{P}\hat{K}_t + \mathbf{Q}\hat{Z}_t \quad (3.45)$$

$$\hat{L}_t = \mathbf{R}\hat{K}_t + \mathbf{S}\hat{Z}_t, \quad (3.46)$$

sustituyendo (3.45) y (3.46) en (3.42), (3.43) y (3.44) se obtiene:

$$\mathbf{A}(\mathbf{P}\hat{K}_t + \mathbf{Q}\hat{Z}_t) + \mathbf{B}\hat{K}_t + \mathbf{C}(\mathbf{R}\hat{K}_t + \mathbf{S}\hat{Z}_t) = 0 \quad (3.47)$$

$$\mathbf{G}(\mathbf{P}\hat{K}_t + \mathbf{Q}\hat{Z}_t) + \mathbf{H}\hat{K}_t + \mathbf{J}(\mathbf{R}(\mathbf{P}\hat{K}_t + \mathbf{Q}\hat{Z}_t) + \mathbf{S}\hat{Z}_t) + \mathbf{K}(\mathbf{R}\hat{K}_t + \mathbf{S}\hat{Z}_t) + \quad (3.48)$$

$$(\mathbf{LN} + \mathbf{M})\hat{Z}_t = 0 \quad (3.49)$$

$$E_t(\hat{Z}_{t+1}) = \mathbf{N}\hat{Z}_t. \quad (3.50)$$

Se tiene un sistema con igual cantidad de ecuaciones con expectativas y de variables de estado endógeno (2 ecuaciones con expectativas y 2 variables de estado endógeno). En el modelo básico de crecimiento, estas ecuaciones necesitan mantenerse para cualquier valor de K_t o Z_t , por tanto es necesario igualar cada coeficiente a cero en K_t y en Z_t , lo que resulta el siguiente sistema de ecuaciones:

$$\mathbf{AP} + \mathbf{B} + \mathbf{CR} = 0 \quad (3.51)$$

$$\mathbf{GP} + \mathbf{H} + \mathbf{JRP} + \mathbf{KR} = 0 \quad (3.52)$$

$$\mathbf{AQ} + \mathbf{CS} + \mathbf{D} = 0 \quad (3.53)$$

$$(\mathbf{G} + \mathbf{JR})\mathbf{Q} + \mathbf{JSN} + \mathbf{KS} + \mathbf{LN} + \mathbf{M} = 0, \quad (3.54)$$

de donde se obtiene:

$$P = -\frac{1}{2} \left[-\left(\frac{B}{A} + \frac{K}{J} - \frac{GC}{JA}\right) \pm \sqrt{\left(\frac{B}{A} + \frac{K}{J} - \frac{GC}{JA}\right)^2 - 4\left(\frac{KB - HC}{JA}\right)} \right] \quad (3.55)$$

$$R = -\frac{1}{C}(AP + B) \quad (3.56)$$

$$Q = \frac{-D(JN + K) + CLN + CM}{AJN + AK - CG - CJR} \quad (3.57)$$

$$S = \frac{-ALN - AM + DG + DJR}{AJN + AK - CG - CJR}. \quad (3.58)$$

A partir de estas soluciones podemos reescribir el modelo básico de crecimiento en forma espacio-estado, y luego usar el Filtro de Kalman para evaluar la verosimilitud.

§3.2. Modelo Básico de Crecimiento en forma de Espacio-Estado.

Consideremos que los datos del modelo están representados en escala logarítmica y están sujetos a un error de medida $V_t = (v_{1,t}, v_{2,t}, v_{3,t}) \sim N(0, \Lambda)$, con $\Lambda = \text{diag}(\sigma_1^2, \sigma_2^2, \sigma_3^2)$. Observese que por la ley de equilibrio recursivo se tienen las ecuaciones (3.45) y (3.46). Por otra parte, tomando logaritmo en la función de producción

$$\begin{aligned} Y_t &= Z_t K_t^\alpha L_t^{1-\alpha} \\ \log(Y_t) &= \log Z_t + \alpha \log K_t + (1 - \alpha) \log L_t \\ \hat{Y}_t &= \hat{Z}_t + \alpha \hat{K}_t + (1 - \alpha) \hat{L}_t, \end{aligned}$$

sustituyendo la ecuación (3.46) se obtiene:

$$\hat{Y}_t = [\alpha + (1 - \alpha)\mathbf{R}]\hat{K}_t + [1 + (1 - \alpha)\mathbf{S}]\hat{Z}_t. \quad (3.59)$$

Finalmente, sustituyendo (3.46) en (3.39) se tiene:

$$\hat{C}_t = (\alpha - \alpha_5\mathbf{R})\hat{K}_t + (1 - \alpha_5\mathbf{S})\hat{Z}_t. \quad (3.60)$$

Por consiguiente, tenemos el siguiente sistema de ecuaciones:

$$\hat{K}_{t+1} = \mathbf{P}\hat{K}_t + \mathbf{Q}\hat{Z}_t \quad (3.61)$$

$$\hat{Z}_{t+1} = \rho\hat{Z}_t + \omega_t \quad (3.62)$$

$$\hat{L}_t = \mathbf{R}\hat{K}_t + \mathbf{S}\hat{Z}_t \quad (3.63)$$

$$\hat{Y}_t = [\alpha + (1 - \alpha)\mathbf{R}]\hat{K}_t + [1 + (1 - \alpha)\mathbf{S}]\hat{Z}_t \quad (3.64)$$

$$\hat{C}_t = (\alpha - \alpha_5\mathbf{R})\hat{K}_t + (1 - \alpha_5\mathbf{S})\hat{Z}_t, \quad (3.65)$$

el cual podemos reescribir como:

Ecuación de observaciones

$$\begin{bmatrix} \log(Y_t) \\ \log(L_t) \\ \log(C_t) \end{bmatrix} = \begin{bmatrix} \alpha + (1 - \alpha)\mathbf{R} & 1 + (1 - \alpha)\mathbf{S} \\ \mathbf{R} & \mathbf{S} \\ \alpha - \alpha_5\mathbf{R} & 1 - \alpha_5\mathbf{S} \end{bmatrix} \begin{bmatrix} \hat{K}_t \\ \hat{Z}_t \end{bmatrix} + \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix}.$$

Ecuación de Estado

$$\begin{bmatrix} \hat{K}_{t+1} \\ \hat{Z}_{t+1} \end{bmatrix} = \begin{bmatrix} \mathbf{P} & \mathbf{Q} \\ 0 & \rho \end{bmatrix} \begin{bmatrix} \hat{K}_t \\ \hat{Z}_t \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \omega_t.$$

§3.3. Filtro de Kalman

El filtro de Kalman es un conjunto de ecuaciones matemáticas que proveen una solución recursiva eficiente del método de mínimos cuadrados. Esta solución permite calcular un estimador lineal, insesgado y óptimo del estado de un proceso en cada momento del tiempo con base en la información disponible en el momento $t - 1$, y actualizar, con la información adicional disponible en el momento t , dichas estimaciones. Este filtro es el principal algoritmo para estimar sistemas dinámicos especificados en la forma de espacio-estado, los cuales tienen muchas aplicaciones econométricas de interés [7].

La representación espacio- estado es esencialmente una notación conveniente para la estimación de modelos estocásticos donde se asumen errores en la medición del sistema, lo que permite abordar el manejo de un amplio rango de modelos de series de tiempo. Entre los usos particulares se encuentra el modelado de componentes no observables y parámetros que cambian en el tiempo, así como la representación de modelos por máxima verosimilitud. El procedimiento de estimación del filtro es el siguiente: el modelo es formulado en su forma espacio- estado y para un conjunto inicial de parámetros dados, los errores de predicción del modelo son generados por el filtro, estos son utilizados para evaluar recursivamente la función de verosimilitud hasta maximizarla. La derivación del Filtro de Kalman se desarrolló en el contexto de modelos espacio- estado, en donde el núcleo es la estimación por medio de mínimos cuadrados recursivos.

Para un modelo lineal dinámico se considera la siguiente ecuación, llamada ecuación de estado:

$$x_t = \Phi x_{t-1} + w_t, \quad (3.66)$$

donde Φ es una matriz de orden $p \times p$ llamada matriz de transición de estados y w_t de $p \times 1$ es el vector de ruido asociado al modelo del sistema, son independientes e idénticamente distribuidos como vectores normales con media cero y matriz de covarianza Q .

En el comienzo del proceso se considera el vector normal x_0 que tiene media μ_0 y matriz de covarianza Σ_0 , de dimensión $p \times p$.

Como no es posible observar el vector estado x_t directamente, entonces mediante lo que llamaremos la ecuación de observaciones podemos representar mediante una función lineal la información de la variable de estado más un cierto ruido, por lo se sigue:

$$y_t = A_t x_t + v_t, \quad (3.67)$$

donde A_t es una matriz de observación o medida de dimensión $q \times p$ y v_t es el ruido aditivo de la observación y es asumido como blanco o gaussiano con matriz de covarianza R de dimensión $q \times q$.

En la práctica el objetivo del análisis que involucre el modelo espacio- estado como se definió en (3.66) y (3.67) será generar estimadores para los vectores no observados x_t , dada la data $Y_s = y_1, \dots, y_s$ a tiempo s . Este tiempo nos lleva a considerar tres tipos de clasificaciones del problema:

- Cuando $s < t$, el problema es llamado de Pronóstico o Predicción.
- Cuando $s = t$, el problema es llamado Filtrado,
- Cuando $s > t$, el problema es denominado Suavizado.

Para desarrollar la teoría pertinente a los Filtros consideraremos las siguientes definiciones:

$$x_t^s = E(x_t | Y_s) \quad (3.68)$$

y

$$P_{t_1, t_2}^s = E(x_{t_1} - x_{t_1}^s)(x_{t_2} - x_{t_2}^s)'. \quad (3.69)$$

Cuando $t_1 = t_2 = t$ y asumimos el proceso como gaussiano, (3.69) se convierte en:

$$P_{t_1, t_2}^s = E\{(x_{t_1} - x_{t_1}^s)(x_{t_2} - x_{t_2}^s)' | Y_s\}. \quad (3.70)$$

El filtro de Kalman genera las ecuaciones de filtro y predicción, el nombre de filtro se deriva de que x_t^t es un filtro lineal de las observaciones y_1, \dots, y_t , esto es

$$x_t^t = \sum_{s=1}^t B_s y_s,$$

para matrices B_s de dimensión $p \times q$ elegidas adecuadamente.

Una ventaja del filtro es que se especifica como actualizar los estados de x_{t-1}^{t-1} hasta x_t^t , a partir de la observación y_t sin tener que procesar todo el conjunto de datos y_1, \dots, y_t .

§3.3.1. Propiedades: Filtro de Kalman

Para el modelo lineal dinámico en forma espacio- estado dado por las siguientes ecuaciones,

$$x_t = \Phi x_{t-1} + w_t, \quad y_t = A_t x_t + v_t,$$

donde Y es de $p \times r$ y Γ es de $q \times r$, con condiciones iniciales $x_0^0 = \mu_0$ y $P_0^0 = \Sigma_0$, para $t = 1, \dots, n$, se tiene que:

$$x_t^{t-1} = \Phi x_{t-1}^{t-1} + Y u_t, \tag{3.71}$$

$$P_t^{t-1} = \Phi P_{t-1}^{t-1} \Phi' + Q, \tag{3.72}$$

con

$$x_t^t = x_t^{t-1} + K_t (y_t - A x_t^{t-1} - \Gamma u_t), \tag{3.73}$$

$$P_t^t = [I - K_t A_t] P_t^{t-1}, \tag{3.74}$$

donde,

$$K_t = P_t^{t-1} A_t' [A_t P_t^{t-1} A_t' + R]^{-1} \tag{3.75}$$

es llamado ganancia de Kalman. La predicción para $t > n$, se logrado vía (3.73) y (3.74) con condiciones iniciales x_n^n y P_n^n

Demostración:

De las desviaciones de (3.72) y (3.73) y realizando algunos cálculos, se tiene

$$\begin{aligned}
 x_t^{t-1} &= E(x_t|Y_{t-1}) \\
 &= E(\Phi x_{t-1} + Y u_t + w_t|Y_{t-1}) \\
 &= \Phi x_{t-1}^{t-1} + Y u_t.
 \end{aligned}$$

Por otro lado;

$$\begin{aligned}
 P_t^{t-1} &= E\{(x_t - x_t^{t-1})(x_t - x_t^{t-1})'\} \\
 &= E\{[\Phi(x_{t-1} - x_{t-1}^{t-1}) + w_t][\Phi(x_{t-1} - x_{t-1}^{t-1}) + w_t]'\} \\
 &= \Phi P_{t-1}^{t-1} \Phi' + Q.
 \end{aligned}$$

Para obtener la ecuación (3.73), primero se definen las innovaciones como:

$$\epsilon_t = y_t - E(y_t|Y_{t-1}) = y_t - A_t x_t^{t-1} - \Gamma u_t, \quad (3.76)$$

para $t = 1, \dots, n$. Note que, $E(\epsilon_t) = 0$.

Además de las definiciones y propiedades de la covarianza, y de la observación:

$$\Sigma_t =^{def} \text{Var}(\epsilon_t) = \text{Var}[A_t(x_t - x_t^{t-1}) + v_t] = A_t P_t^{t-1} A_t' + R, \quad (3.77)$$

donde $E(\epsilon_t y_s')$ para $s < t$.

Ya que la sucesión de innovaciones es un proceso Gaussiano, entonces las innovaciones son independientes de las observaciones pasadas. Más aún, la covarianza condicional entre x_t y ϵ_t dado Y_{t-1} , es:

$$\begin{aligned}
 \text{Cov}(x_t, \epsilon_t | Y_{t-1}) &= \text{Cov}(x_t, y_t - A_t x_t^{t-1} - \Gamma u_t | Y_{t-1}) \\
 &= \text{Cov}(x_t x_t^{t-1}, y_t - A_t x_t^{t-1} - \Gamma u_t | Y_{t-1}) \\
 &= \text{Cov}[x_t x_t^{t-1}, A_t(x_t - x_t^{t-1}) + v_t] \\
 &= P_t^{t-1} A_t'.
 \end{aligned} \quad (3.78)$$

Usando estos resultados se tiene que la distribución condicional conjunta de x_t y ϵ_t dado Y_{t-1} es normal;

$$X \sim N(\mu, \sigma^2),$$

con

$$X = \begin{bmatrix} x_t \\ \epsilon_t \end{bmatrix}.$$

Además con,

$$\begin{pmatrix} x_t \\ \epsilon_t \end{pmatrix} | Y_{t-1} \sim N \left(\begin{bmatrix} E(x_t | Y_{t-1}) \\ E(\epsilon_t | Y_{t-1}) \end{bmatrix}, \begin{bmatrix} \mathbf{V}(x_t | Y_{t-1}) & \mathbf{C}(x_t | Y_{t-1}) \\ \mathbf{C}(\epsilon_t | Y_{t-1}) & \mathbf{V}(\epsilon_t | Y_{t-1}) \end{bmatrix} \right), \quad (3.79)$$

se tiene que:

$$\begin{pmatrix} x_t \\ \epsilon_t \end{pmatrix} | Y_{t-1} \sim N \left(\begin{bmatrix} x_t^{t-1} \\ 0 \end{bmatrix}, \begin{bmatrix} P_t^{t-1} & P_t^{t-1} A_t' \\ A_t P_t^{t-1} & \Sigma_t \end{bmatrix} \right). \quad (3.80)$$

Por otra parte,

$$x_t^t = E(x_t | y_1, \dots, y_{t-1}, y_t) = E(x_t | Y_{t-1}, \epsilon_t) = x_t^{t-1} + K_t \epsilon_t, \quad (3.81)$$

de donde,

$$K_t = P_t^{t-1} A_t' \Sigma_t^{-1} = P_t^{t-1} A_t' (A_t P_t^{t-1} A_t' + R)^{-1}. \quad (3.82)$$

La evolución de P_t^t es fácilmente calculada de (3.82) como

$$P_t^t = \text{cov}(x_t | Y_{t-1}, \epsilon_t) = P_t^{t-1} - P_t^{t-1} A_t' \Sigma_t^{-1} A_t P_t^{t-1}.$$

§3.4. Estimación de Máxima Verosimilitud

El estimador de los parámetros especificados en el modelo espacio-estado, (3.66) y (3.67), está absolutamente implicado. Sea $\Theta = \{\mu_0, \Sigma_0, \Phi, Q, R, Y, \Gamma\}$ el vector de parámetros que contiene los elementos de la media inicial y la covarianza μ_0 y Σ_0 , la matriz de transición Φ , y las matrices de estado y observación Q y R y las matrices

de coeficientes de entrada, Y y Γ , usaremos la máxima verosimilitud bajo la suposición de que el estado inicial es normal, $x_0 \sim N(\mu_0, \Sigma_0)$, y los errores w_1, \dots, w_n y v_1, \dots, v_n son vectores variables conjuntamente normal y no correlacionados. Asumiremos, por simplicidad que, $\{w_t\}$ y $\{v_t\}$ son no correlacionado.

La verosimilitud es calculada usando las innovaciones $\epsilon_1, \epsilon_2, \dots, \epsilon_n$, definidos por (3.78),

$$\epsilon_t = y_t - A_t x_t^{t-1} - \Gamma u_t.$$

La forma de actualizar la función de verosimilitud es obtenido usando un argumento similar, los procesos para notar las innovaciones son vectores aleatorios Gaussianos independientes con media cero y, como demostramos en (3.77), la matriz de covarianza

$$\Sigma_t = A_t P_t^{t-1} A_t' + R. \quad (3.83)$$

Por lo tanto, ignorando la constante, podemos escribir la verosimilitud, $L_Y(\Theta)$, como

$$-\ln L_Y(\Theta) = \frac{1}{2} \sum_{t=1}^n \log |\Sigma_t(\Theta)| + \frac{1}{2} \sum_{t=1}^n \epsilon_t(\Theta)' \Sigma_t(\Theta)^{-1} \epsilon_t(\Theta), \quad (3.84)$$

donde se destaca la independencia de las innovaciones sobre los parámetros Θ . Por supuesto, (3.84) es altamente no lineal y una función complicada de los parámetros desconocidos. El procedimiento usual es un x_0 fijo y entonces se convierte en un conjunto de recursiones para la función de log verosimilitud y sus primeras dos derivadas.

Note que en las ecuación espacio-estado se supone que:

$$\omega_t \sim N(0, W_t) \quad y \quad v_t \sim N(0, V_t).$$

La función del Filtro de Kalman permite recursivamente construir un pronóstico lineal de y_t , llamado $y_{t|t-1}$, basado en las observaciones anteriores. Además, esto

implica que la diferencia de y_t y su predicción $y_{t|t-1}$ se distribuye normalmente con media cero y varianza Q_t . La salida del filtro de Kalman puede ser usado para calcular la función de verosimilitud dada por:

$$L(y^T|\gamma) = \prod_{t=1}^T \frac{1}{\sqrt{(2\pi)^n |Q_t}} \exp \left\{ -\frac{(y_t - y_{t|t-1})' Q_t^{-1} (y_t - y_{t|t-1})}{2} \right\}, \quad (3.85)$$

y la log-verosimilitud está dada por:

$$\ln L(y^T|\gamma) = -\frac{nT}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^T \ln |Q_t(\gamma)| - \frac{1}{2} \sum_{t=1}^T e_t'(\gamma) Q_t^{-1}(\gamma) e_t(\gamma) \quad (3.86)$$

donde, por (52),

$$\begin{aligned} e_t = y_t - y_{t|t-1} &= y_t - F_t E(\theta_t | D_{t-1}) - d_t \\ &= F_t \theta_t + d_t + v_t - F_t E(\theta_t | D_{t-1}) - d_t \end{aligned} \quad (3.87)$$

$$= F_t (\theta_t - E(\theta_t | D_{t-1})) + v_t. \quad (3.88)$$

§3.4.1. Función de Verosimilitud asociada al Modelo

La verosimilitud para el modelo básico de crecimiento, queda como:

$$\log L = -\frac{3T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T \log |Q_t| - \frac{1}{2} \sum_{t=1}^T e_t' Q_t^{-1} e_t \quad (3.89)$$

donde,

$$e_t = y_t - y_{t|t-1} = y_t - F_t E(\theta_t | D_{t-1}),$$

$$Q_t = E(e_t e_t') = F_t \text{Var}(\theta_t | D_{t-1}) F_t' + V_t, \quad (3.90)$$

e_t representa el error de predicción y Q_t es su matriz del error cuadrático medio.

El procedimiento usual para lograr nuestro objetivo es fijar un x_0 y entonces desarrollar un conjunto de recursiones para la función de verosimilitud y su primera y

segunda derivada. Entonces con un algoritmo de SPSA que puede ser usado sucesivamente para actualizar los valores de los parámetros hasta lograr minimizar la verosimilitud negativa.

§3.5. Algoritmo Metropolis-Hasting

El algoritmo Metropolis- Hasting es un poderoso método de Cadenas de Markov, este algoritmo propone una distribución $q(x, y)$ que sea fácil de muestrear, donde $\int q(x, y)dy = 1$; es decir, $q(x, y)$ es una función de densidad de probabilidad. Esta distribución se conoce como probabilidad de transición; es decir, la probabilidad de moverse de x a y . Cuando un proceso está en el punto x la densidad genera un valor y de $q(x, y)$. Ahora:

Si $q(x, y)$, satisface la siguiente condición, llamada condición de reversibilidad:

$$\pi(x)q(x, y) = \pi(y)q(y, x), \quad (3.91)$$

se finaliza la búsqueda.

Lo más frecuente es que el caso anterior no ocurra, y por tanto para algún x, y ,

$$\pi(x)q(x, y) > \pi(y)q(y, x). \quad (3.92)$$

En este caso, se dice que el proceso se mueve de x a y con más frecuencia que de y a x . Un camino conveniente es reducir el número de movimientos de x a y introduciendo una probabilidad:

$$\alpha(x, y) < 1, \quad (3.93)$$

llamada probabilidad de movimiento o de transición; es decir, la probabilidad de que el movimiento se realice. Si éste movimiento no se realiza entonces el proceso retorna x como un valor de la distribución. Entonces la transición de x a y ($x \neq y$) es realizada de acuerdo:

$$P_{MH}(x, y) \equiv q(x, y)\alpha(x, y), \quad x \neq y, \quad (3.94)$$

donde $\alpha(x, y)$ es determinada. Considerando de nuevo (3.94), se tiene que el movimiento de y a x no ocurre muy a menudo. Por lo tanto, se debe definir $\alpha(x, y)$ tan grande como sea posible, debido a que es una probabilidad, su límite superior es 1. Ahora, la probabilidad de movimiento $\alpha(x, y)$ es determinada requiriendo que P_{MH} satisfaga la condición de reversibilidad, porque entonces:

$$\pi(x)q(x, y)\alpha(x, y) = \pi(y)q(y, x)\alpha(y, x) \quad (3.95)$$

$$= \pi(y)q(y, x), \quad (3.96)$$

lo que implica:

$$\alpha(x, y) = \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)} \quad (3.97)$$

es decir, la transición P_{MH} satisface la condición de reversibilidad. Si la desigualdad se invierte entonces se coloca $\alpha(y, x) = 1$ y se halla $\alpha(y, x)$ de la misma manera. De ésta forma, se ha probado que para ser reversible la transición P_{MH} , la probabilidad de movimiento debe ser:

$$\alpha(x, y) = \begin{cases} \min\left[\frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}, 1\right], & \text{Si } \pi(x)q(x, y) > 0; \\ 1, & \text{En cualquier otro caso.} \end{cases}$$

El algoritmo genera una cadena $\{\gamma_i\}_{i=1}^M$ de $\pi_j(\gamma|y^T)$ de la siguiente manera:

Paso 1. Se considera in valor inicial, $\tilde{\theta}$.

Paso 2. Sea $\tilde{\Sigma}$, la inversa de la matriz Hessiana calculada en la moda $\tilde{\theta}$.

Paso 3. Generamos θ_0 de $N(\tilde{\theta}, c_0^2\tilde{\Sigma})$.

Paso 4. Para $s = 1, \dots, n$, tomar ϑ desde la distribución propuesta $N(\theta_{s-1}, c^2\tilde{\Sigma})$. Si la actualización de θ_{s-1} es aceptada ($\theta_s = \vartheta$) con probabilidad $\min\{1, r(\theta_{s-1}, \vartheta|Y)\}$ y en cualquier otro caso se rechaza ($\theta_s = \theta_{s-1}$). Aquí

$$r(\theta_{s-1}, \vartheta|Y) = \frac{L(\vartheta|Y)p(\vartheta)}{L(\theta_{s-1}|Y)p(\theta_{s-1})}$$

Paso 5. Aproximar el valor esperado de el postirio de la función $h(\theta)$ por $\frac{1}{n} \sum_{s=1}^n h(\theta_s)$.

Realizando una modificación en el algoritmo anterior, de tal forma que realice una actualización considerando la dirección dada por el algoritmo de SPSA descrito en el capítulo 2. Este algoritmo lo denominaremos MH-SPSA y tiene la siguiente estructura:

Paso 1. Se considera in valor inicial, $\tilde{\theta}$.

Paso 2. Sea $\tilde{\Sigma}$, se determina la inversa de la matriz Hessiana calculada como en la ecuación (2.25)

Paso 3. Para $s = 1, \dots, n$, tomar ϑ desde,

$$\vartheta_s = \vartheta_{i-1} - a_{s-1} H_{s-1}^{-1} \hat{g}_{s-1}(\vartheta_{i-1}),$$

donde a_s es una sucesión de constantes positivas tal que, $0 < \sum_{n=1}^{\infty} a_n^2 = A < \infty$.

Si la actualización de θ_{s-1} es aceptada ($\theta_s = \vartheta$) con probabilidad $\min\{1, r(\theta_{s-1}, \vartheta|Y)\}$ y en cualquier otro caso se rechaza ($\theta_s = \theta_{s-1}$).

Aquí

$$r(\theta_{s-1}, \vartheta|Y) = \frac{L(\vartheta|Y)p(\vartheta)}{L(\theta_{s-1}|Y)p(\theta_{s-1})}$$

Paso 4. Aproximar el valor esperado del postirio de la función $h(\theta)$ por $\frac{1}{n} \sum_{s=1}^n h(\theta_s)$.

La dificultad en resolver este algoritmo radica en que no se conoce con precisión la función, puesto que esta contaminada con ruido. Así, la cadena generada por el algoritmo es construida explotando la estructura especial del problema y no es posible establecer un criterio de convergencia de forma analítica[1]. El camino sería construir una Cadena de Markov irredicible, no periódica tal que su distribución de probabilidad invariante sea máxima en ϑ^* , dicha convergencia de la cadena generada por el algoritmo no se ha logrado demostrar.

CAPÍTULO 4

RESULTADOS NUMÉRICOS

Dado el modelo básico de crecimiento estocástico cuyas especificaciones fueron dadas en el Capítulo 3, para poder realizar inferencia Bayessiana se deben especificar las distribuciones a priori para los parámetros del modelo, los cuales vienen dados por:

$$\gamma = \{\theta, \rho, \alpha, \tau, \delta, \beta, \sigma_\varepsilon, \sigma_1, \sigma_2, \sigma_3\}.$$

Por lo tanto, la previa del modelo considerando independencia entre los parámetros es:

$$\pi(\gamma) = \pi(\theta, \rho, \alpha, \tau, \delta, \beta, \sigma_\varepsilon, \sigma_1, \sigma_2, \sigma_3) \quad (4.1)$$

$$\propto \pi(\theta)\pi(\rho)\pi(\alpha)\pi(\tau)\pi(\delta)\pi(\beta)\pi(\sigma_\varepsilon)\pi(\sigma_1)\pi(\sigma_2)\pi(\sigma_3). \quad (4.2)$$

Es claro que se realizará una estimación de la función de verosimilitud a partir de la construcción de la forma espacio-estado del modelo, la cual esta caracterizada mediante la siguiente expresión:

$$\log(L) = -\frac{3T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^T \log|Q_t| - \frac{1}{2}\sum_{t=1}^T e_t'Q_t^{-1}e_t, \quad (4.3)$$

donde e_t representa el error de predicción y Q_t es su matriz de error cuadrático medio.

Se busca realizar la incorporación de la función de verosimilitud en un algoritmo de estimación bayessiana. Esto debido a que la aproximación Bayessiana es una herramienta principal para determinar la distribución a posteriori de los parámetros

dada la data. La densidad a posteriori es proporcional a los tiempos de la verosimilitud de la priori. Por lo tanto, necesitamos estimar priori sobre los parámetros, $\pi(\gamma)$ y evaluar en la función de verosimilitud.

Estos parámetros son evaluados en la función de verosimilitud, considerando que se tiene un modelo de aproximación lineal. Usaremos el algoritmo de Metropolis-Hasting para aproximar sobre la verosimilitud. Este algoritmo nos proporciona una cadena $\{\gamma_i\}_{i=1}^M$, la cual converge a γ^* .

Daremos un soporte a priori para los 10 parámetros, sujeto a algunas restricciones de borde para hacer los priori apropiadas, esto no es más que considerar previas suaves para los parámetros sujetos solamente a algunas restricciones para garantizar que las priori sean propias. Esta elección es motivada por dos consideraciones. Primero, puesto que vamos a estimar nuestro modelo usando data "artificial" generado por algún valor de γ^* , no queremos basar nuestros resultados en cualquier alternativa por nuestra elección de priori. Segundo, con el plano de priori, la posteriori es proporcional a la función de verosimilitud (excepto para una cantidad pequeña de el soporte de las cotas de las priori). Como una consecuencia, nuestro experimento puede ser interpretado como un ejercicio clásico, en el cual la función de verosimilitud es la estimación del máxima verosimilitud.

§4.1. Parámetros del Modelo Macroeconómico

La interpretación de los parámetros desde el punto de vista económico es el siguiente, el parámetro de mano de obra gubernamental, θ , sigue una distribución uniforme entre 0 y 1. Esta restricción impuesta es solo para una utilidad marginal positiva de ocio. La persistencia del shock tecnológico, ρ , también sigue una distribución uniforme entre 0 y 1, esta región implica una distribución estacionaria de las variables del modelo con una cota inferior sobre la no persistencia. El parámetro τ sigue una distribución uniforme entre 0 y 100. Ésta elección abarca todo el rango de los estimadores empíricos de los parámetros y solamente toma en cuenta los riesgos

afectos y los riesgos no afectos que podrían predecir diferencias en las tasas de interés de grandes magnitudes que se pueden observar. La priori para el parámetro tecnológico, α , es uniforme entre 0 y 1, esto incluye todos los valores para los cuales la productividad marginal de capital y el trabajo son positivas. La priori sobre los rangos de la tasa de depreciación, δ esta en un rango entre 0 y 0.05, cubriendo todas las estimaciones de la contabilidad nacional de la depreciación trimestral. El factor de descuento, β , se permite variar entre 0.75 y 1, implicando que la tasa de interés anual del estado es estable y esta entre 0 y 316%. La desviación estándar de las innovaciones de productividad, σ_ε , sigue una distribución uniforme entre 0 y 0.1, una cota de 15 veces más alto que las estimaciones usuales. También elegimos este anterior para tres desviaciones estándar de los errores de la media de la ecuación de observación. En la siguiente tabla se muestran las cotas de cada uno de estos parámetros.

Parámetros	Distribución	Hiperparámetros
θ	Uniforme	0,1
ρ	Uniforme	0,1
τ	Uniforme	0,100
α	Uniforme	0,1
δ	Uniforme	0,0.05
β	Uniforme	0,0.75
σ_ε	Uniforme	0,0.1
σ_1	Uniforme	0,0.1
σ_2	Uniforme	0,0.1
σ_3	Uniforme	0,0.1

Las estimaciones fueron realizadas con el algoritmo MH-SPSA descrito en el capítulo 3; que genera una sucesión y lo que perseguimos es determinar un contraste entre el algoritmo Metropolis- Hasting y el algoritmo propuesto.

§4.2. Data para la experimentación

§4.2.1. Experimentaciones con Data Calibrada

Se realizaron estimaciones sobre las observaciones desde el modelo y usaremos estas como data para la estimación. Generaremos data con dos calibraciones diferentes. Primero, tomaremos el experimento lo más realista posible, presentaremos calibración del patrón de prueba del modelo. El factor de descuento $\beta = 0,9896$ considera una tasa de interés anual de 4,27 %. El riesgo de error $\tau = 2$ es una elección común en la literatura. $\theta = 0,357$ se aproxima mucho a la oferta de mano de obra. Reproducimos la parte de trabajo de la renta nacional con $\alpha = 0,4$. La tasa de depreciación $\delta = 0,02$ fija la investigación sobre el cociente de salida $\rho = 0,95$ y $\sigma = 0,007$. Con respecto a la desviación estándar de los errores de las mediciones, el conjunto de ellos es igual 0,01 % de el valor de la salidas del estado estable, 0,35 % de el estado estable del valor de horas y 0,2 %, de el valor del estado estacionario basado en nuestras priori con respecto a la importancia relativa de las mediciones de los errores.

La segunda calibración, la cual llamaremos extrema, sigue los mismos parámetros que la data descrita anteriormente, salvo que en este caso consideraremos $\tau = 50$, esto implica un error relativo de 24.5, además el parámetro $\sigma_e = 0,035$. Este error y la variación de riesgo elevado introducen una ausencia de linealidad fuerte a la economía [10].

Después de hacer distintas simulaciones para tamaños de muestras diferentes y tomando valores iniciales apropiados, con el objetivo de dejar que sean los datos los que determinen el comportamiento del modelo y no las priori. Realizamos experimentación con la data simulada para cada una de las calibraciones antes descritas, aplicamos las priori y evaluamos los algoritmos para la función de verosimilitud. En el algoritmo de MH-SPSA se obtuvo la mejor estimación para 10000 iteraciones, mientras que para el algoritmo de Metropolis- Hasting se realizaron 50000 estimaciones. Dadas estas especificaciones obtuvimos los siguientes resultados, los cuales podemos resumir a continuación, además que estos resultados serán comparados con los obtenido por Rubio en [10].

Estimador a posteriori						
	MH-SPSA		Metropolis		Rubio	
Parámetro	Media	Std	Media	Std	Media	Std
θ	0.3760	0.0167	0.4481	0.0015	0.374	0.06×10^{-3}
ρ	0.9086	0.0060	0.9001	1.02×10^{-4}	0.914	0.24×10^{-3}
τ	4.01	0.0026	3.5998	1.47×10^{-4}	3.536	1.17×10^{-3}
α	0.4167	0.0093	0.3898	2.65×10^{-4}	0.443	0.08×10^{-3}
δ	0.0484	0.0042	0.0220	0.0017	0.030	0.02×10^{-3}
β	0.9530	0.0018	0.9483	9.54×10^{-4}	0.978	0.03×10^{-3}
σ_ε	0.0298	0.0041	0.0075	0.0010	0.011	0.36×10^{-4}
σ_1	0.0397	0.0054	0.0035	0.0015	1.86×10^{-4}	4.64×10^{-7}
σ_2	0.0702	0.0087	0.0132	0.0059	5.55×10^{-4}	2.01×10^{-6}
σ_3	0.0022	0.0034	0.0149	0.0069	2.42×10^{-3}	1.75×10^{-6}

TABLA 4.1: Estimaciones: MH-SPSA versus Metropolis- Hasting con la Data Calibrada 1

En las Tabla 4.1, 4.2 y 4.3 se muestran los resultados obtenidos al realizar los experimentos con la data calibrada 1, data calibrada 2 y data real respectivamente. En estas tablas se realiza un análisis comparativo entre el algoritmo de Metropolis-Hasting y el algoritmo de MH-SPSA, donde se determina la media y desviación estándar (std). Se puede apreciar en estos resultados que las estimaciones realizadas con MH-SPSA son los que más se aproximan a los resultados obtenidos por Rubio en [10].

En la Figura 4.1 y Figura 4.2 se pueden apreciar los gráficos de los 10 parámetros observados para las estimaciones realizadas con Data Calibrada 1. En estas gráficas lo que se busca es apreciar un patrón de convergencia para cada parámetro estimado y que esto indique la convergencia que no es fácil de determinar de forma analítica .

Para el caso de los resultados de las estimaciones obtenidas usando la data calibrada 2 o data extrema como también es llamada, podemos observar las estimaciones resumidas en las Figuras 4.3 y 4.4. En estos casos es importante destacar que como nos referimos a la data considerada como extrema la convergencia es más difícil de observar. Por ultimo, consideramos los gráficos obtenido con las estimaciones

Estimador a posteriori						
	MH-SPSA		Metropolis		Rubio	
Parámetro	Media	Std	Media	Std	Media	Std
θ	0.3706	0.0183	0.4480	0.0015	0.337	0.06×10^{-3}
ρ	0.8919	0.0021	0.8998	1.72×10^{-4}	0.894	0.19×10^{-3}
τ	60.0062	0.0012	60.00	6.57×10^{-5}	67.70	0.10×10^{-1}
α	0.4068	0.0067	0.3898	9.28×10^{-5}	0.346	0.04×10^{-3}
δ	0.0315	0.0016	0.0220	0.0017	0.010	0.02×10^{-4}
β	0.9514	0.0014	0.9482	0.0012	0.996	0.08×10^{-4}
σ_ϵ	0.0319	0.0042	0.0073	7.27×10^{-4}	3.61×10^{-2}	0.09×10^{-4}
σ_1	0.0154	0.0075	0.0075	0.0035	1.72×10^{-4}	0.05×10^{-6}
σ_2	0.0794	0.0092	0.0153	0.0069	9.08×10^{-4}	0.03×10^{-5}
σ_3	0.0025	0.0054	0.0111	0.0046	2.64×10^{-3}	0.04×10^{-5}

TABLA 4.2: Estimaciones: MH-SPSA vs Metropolis- Hasting con la Data Calibrada 2 caso extrema

de la Data Real, se puede observar un mejor patron de convergencia.

Otro característica que podemos apreciar y no menos importante ya que conforma una información vital, es el número de evaluaciones que se realizan sobre la función de verosimilitud. En el caso del algoritmo MH-SPSA, este realiza más evaluaciones la función, 50001, mientras Metropolis- Hasting, realiza 50000 evaluaciones, considerando que con el método de MH-SPSA se realizaron 10000 iteraciones mientras que con Metropolis- Hasting se realizaron 50001. Este incremento en el número de iteraciones el algoritmo de Metropolis Hasting es motivado a que el algoritmo tarda más en arrojar buenos resultado para la estimación[10].

En el caso de las estimaciones relizadas por Jesús Fernández Villaverde and Juan F. Rubio Ramirez [10], para el algoritmo allí propuesto logrará resultados satisfactorios al realizar 60.000 iteraciones, lo cual constituye una desventaja al momento de la implementación. Podemos resaltar que el algoritmo MH-SPSA reduce significativamente el número de iteraciones.

Estimador a posteriori						
Parametro	MH-SPSA		Metropolis		Rubio	
	Media	Std	Media	Std	Media	Std
θ	0.4752	0.0042	0.4522	0.0017	0.423	$\times 10^{-3}$
ρ	0.8620	0.0070	0.8499	1.20×10^{-4}	0.941	0.27×10^{-3}
τ	1.6244	0.0046	1.6002	1.83×10^{-4}	1.965	0.12×10^{-2}
α	0.4895	0.0042	0.4997	1.80×10^{-4}	0.412	0.37×10^{-4}
δ	0.0488	0.0054	0.0173	0.0018	0.019	0.79×10^{-5}
β	0.8950	0.0086	0.8509	7.76×10^{-4}	0.973	0.43×10^{-5}
σ_ε	0.0527	0.0064	0.0108	0.0021	0.009	0.43×10^{-5}
σ_1	0.0578	0.0068	0.0109	0.0055	0.11×10^{-3}	0.18×10^{-6}
σ_2	0.0521	0.0063	0.0138	0.0062	0.83×10^{-2}	0.34×10^{-5}
σ_3	0.0481	0.0075	0.0111	0.0051	0.26×10^{-5}	0.18×10^{-4}

TABLA 4.3: Estimaciones: MH-SPSA versus Metropolis- Hasting considerando data real

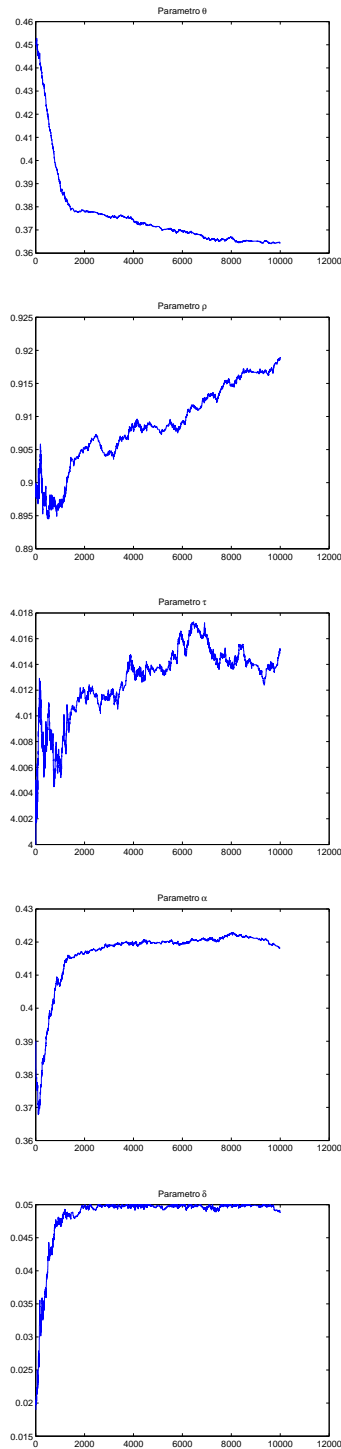
§4.2.2. Experimentaciones con la data real

Ahora estimaremos el modelo neoclásico de macroeconomía con data trimestral. Usaremos salidas verdaderas por capital, las horas de trabajo medias e inversión fija gruesa por capital a partir de 1964 : Q1 a 2003 : Q1. Primero se remueve la tendencia desde la data usando un filtro. De esta manera, no necesitamos modelar explícitamente la presencia de una tendencia y de sus cambios posibles.

Es difícil documentar si las diferencias en las estimaciones del punto que se encuentran resumidos en la Tabla 4.3 son grandes o pequeños. Una posible respuesta se basa en el impacto de las diversas estimaciones en los momentos generados por el modelo. Los macroeconomistas utilizan a menudo estos momentos para evaluar la capacidad del modelo de explicar los datos.

Podemos destacar que dado los resultado obtenidos, el algoritmo MH-SPSA es el que nos provee de un mejor aproximación. Una de las razones de las diferencias entre los algoritmos, es que un factor de descuento más bajo, induce una acumulación más alta de capital.

Estimaciones con MH-SPSA



Estimaciones con Metropolis- Hasting

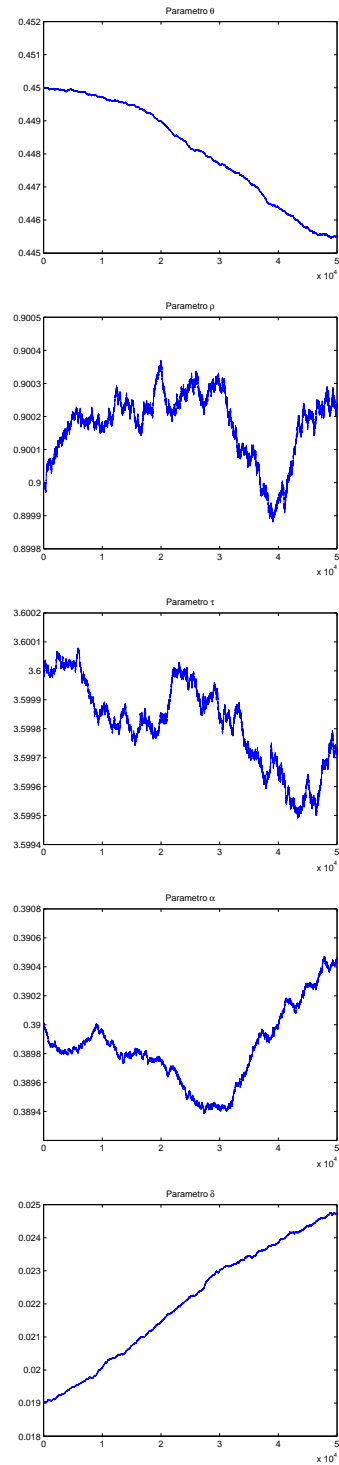
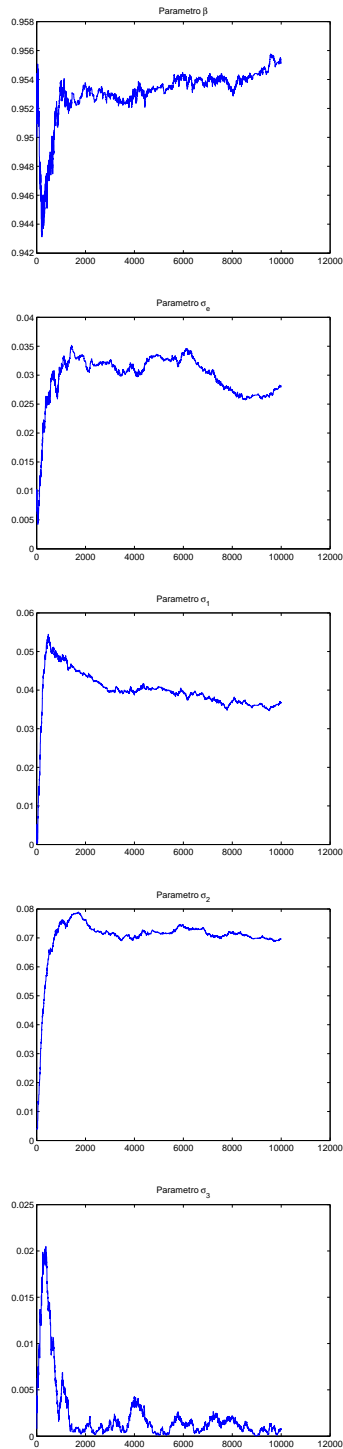


FIGURA 4.1: Estimaciones: MH-SPSA vs Metropolis- Hasting con la Data Calibrada 1

Estimaciones con MH-SPSA



Estimaciones con Metropolis- Hasting

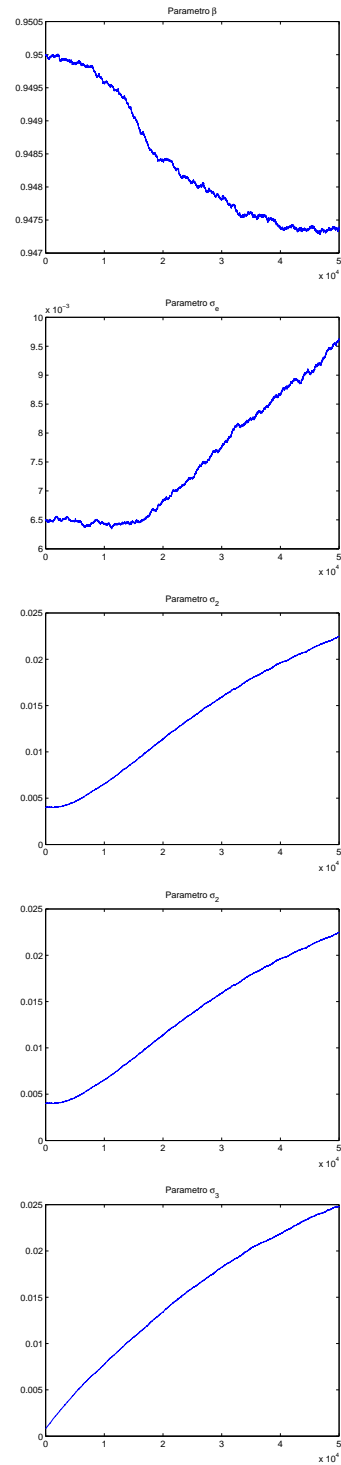
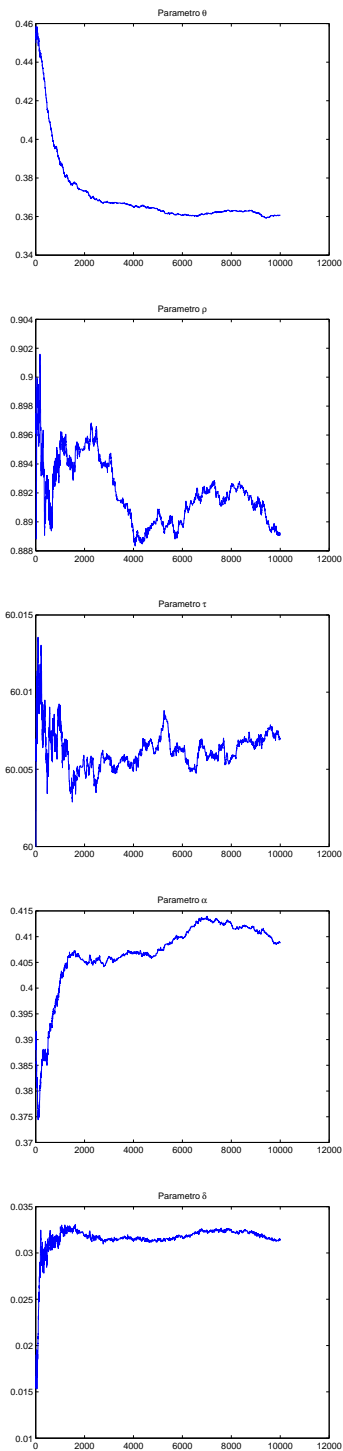


FIGURA 4.2: Estimaciones: MH-SPSA vs Metropolis- Hasting con la Data Calibrada 1

Estimaciones con MH-SPSA



Estimaciones con Metropolis- Hasting

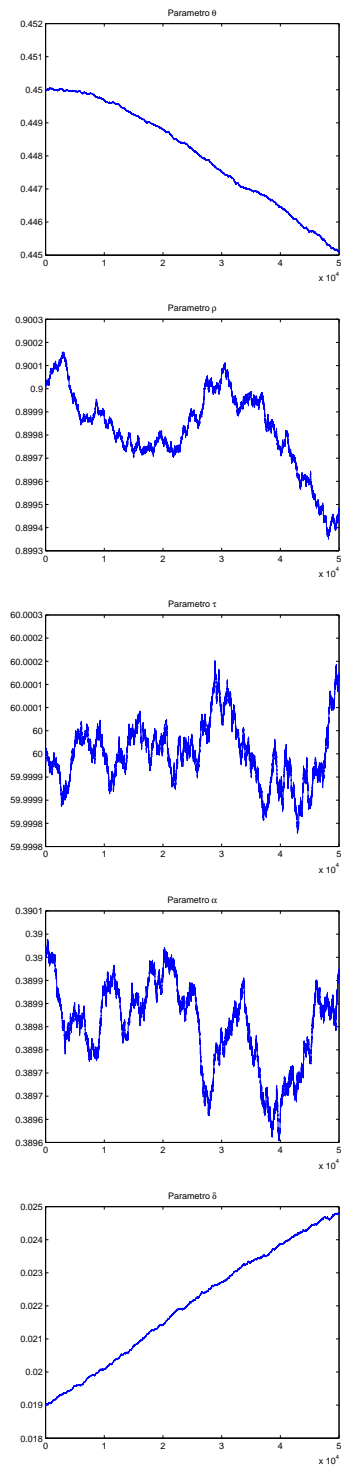
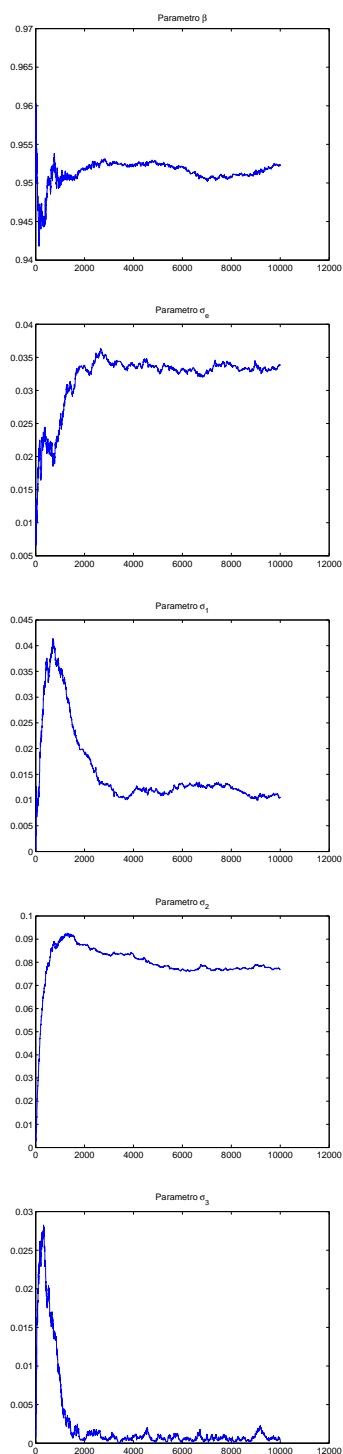


FIGURA 4.3: Estimaciones: MH-SPSA vs Metropolis- Hasting con la Data Calibrada 2

Estimaciones con MH-SPSA



Estimaciones con Metropolis- Hasting

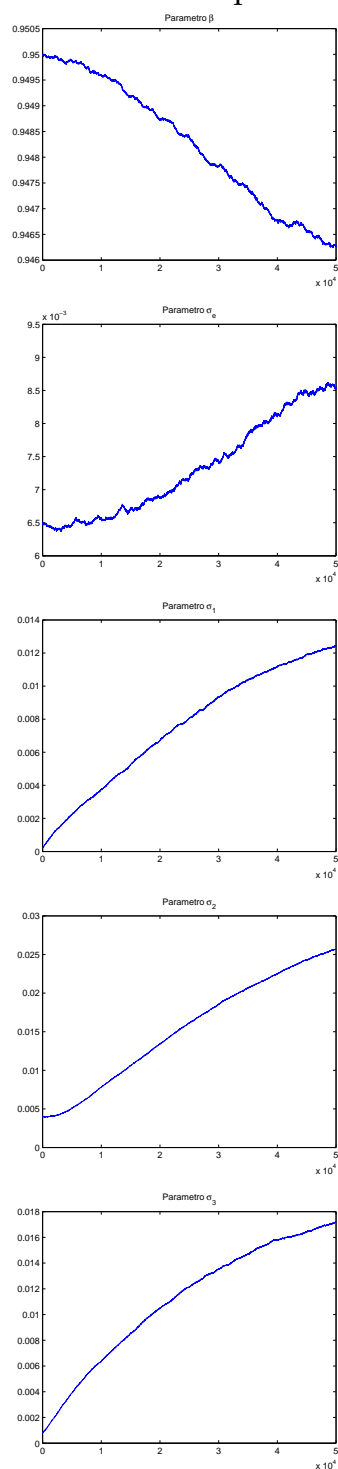
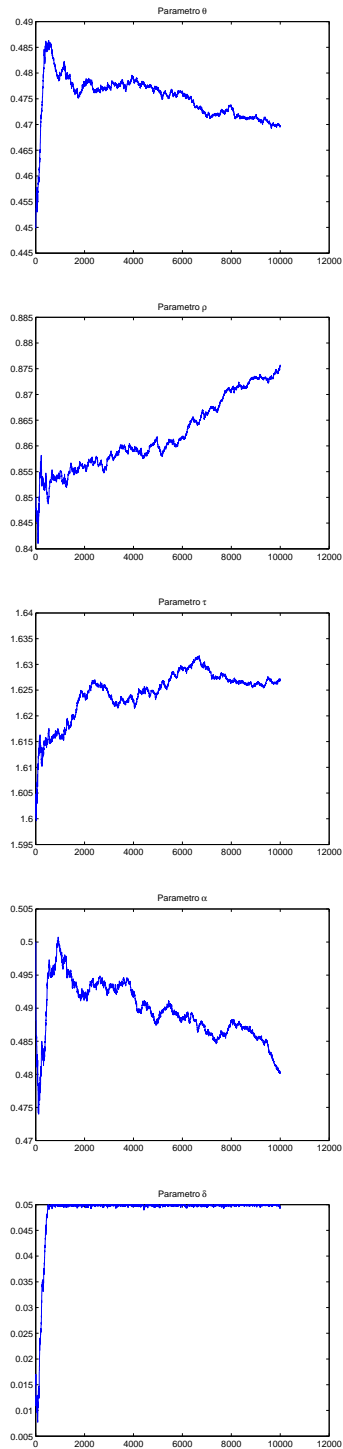


FIGURA 4.4: Estimaciones: MH-SPSA vs Metropolis- Hasting con la Data Calibrada 2

Estimaciones con MH-SPSA



Estimaciones con Metropolis- Hasting

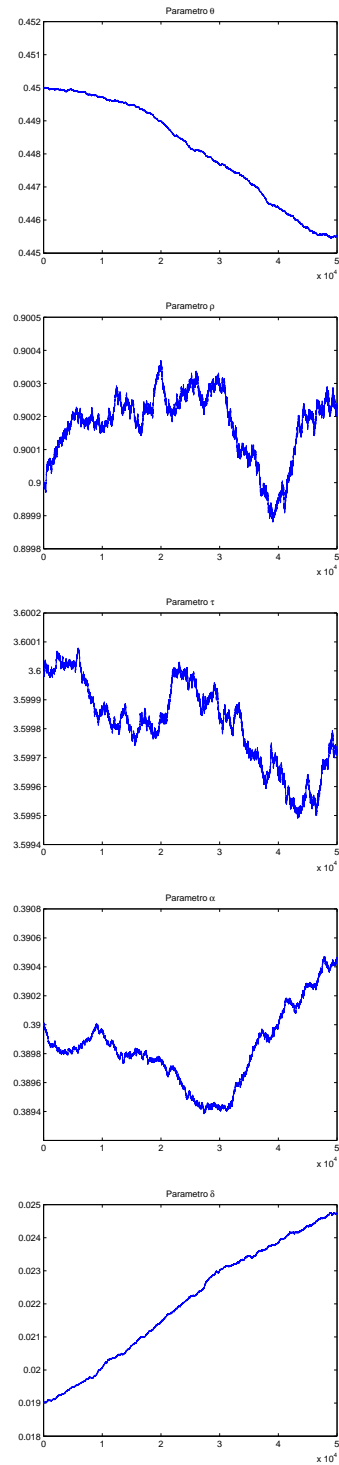
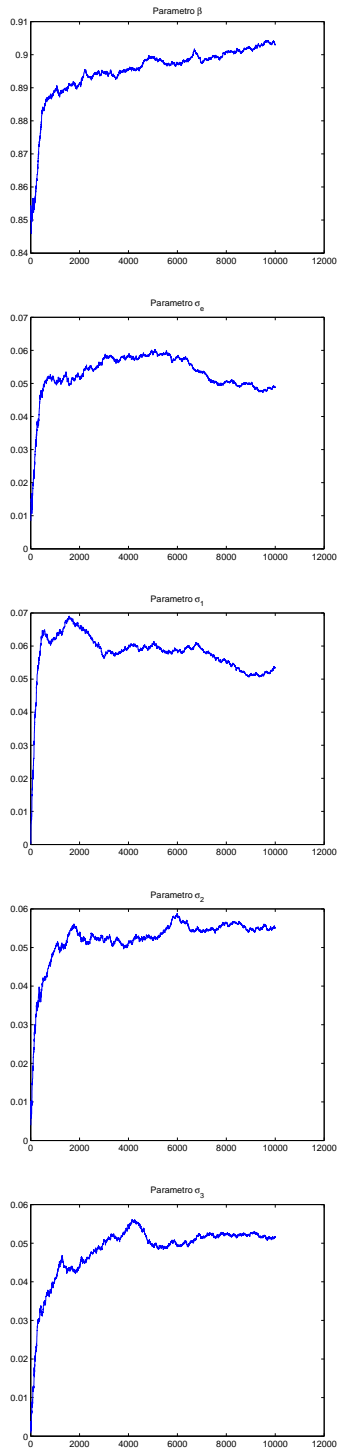


FIGURA 4.5: Estimaciones: MH-SPSA vs Metropolis- Hasting considerando Data Real

Estimaciones con MH-SPSA



Estimaciones con Metropolis- Hasting

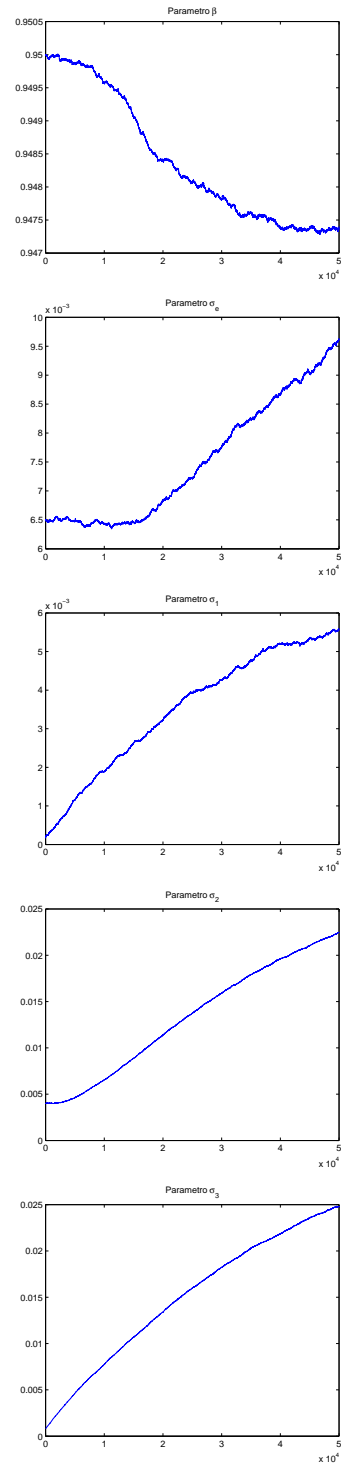


FIGURA 4.6: Estimaciones: MH-SPSA vs Metropolis- Hasting considerando Data Real

CONCLUSIONES

De las experimentaciones realizadas podemos destacar lo siguientes:

- Las estimaciones realizadas con Metropolis- Hasting y MH-SPSA se verifico que MH-SPSA evidencia más eficiencia.
- El algoritmo de MH-SPSA requiere menos iteraciones para llegar a mejores estimaciones.
- El algoritmo de MH-SPSA requiere menos evaluaciones de la función de verosimilitud. El algoritmo de Metropolis- Hasting realiza casi el doble de las evaluaciones de la función.
- Mediante los gráficos de los parámetros estimados podemos apreciar el comportamiento de la convergencia del método.

Es importantte destacar que la teoría de convergencia para el algoritmo MH-SPSA analíticamente no ha sido demostrado, esto debido a la complejidad que existe en las distribuciones, es por ello que en este trabajo sólo logramos revisar con experimentacines la convergencia de forma gráfica y con la información que nos proporcionan las estimaciones realizados. Queda para futuras investigaciones el lograr probar convergencia para este tipo de algoritmos.

REFERENCIAS

- [1] László Gerencsér, Zsuzsanna Vágó Stacy D. Hill, and Zoltán Vincze. Discrete optimization, spsa and markov chain carlo methods. *Proceeding of the 2004 American Control Conference*, pages 3814–3819, 2004.
- [2] Kiefer J. and J. Wolfowitz. Stochastic estimation of a regression function. *Ann. Math. Stat.*, 23:462–466, 1952.
- [3] Kushner H. J. and D. S. Clark. Stochastic approximation methods for constrained and unconstrained systems. *Springer-Verlag. New York*, 1978.
- [4] Blum J. R. Multidimensional stochastic approximation methods. *Ann. Math. Stat.*, 25:737–744, 2000.
- [5] Horn RA and Johnson CR. Matrix analysis. *Cambridge University Press*, 1985.
- [6] Herbert Robbins and Sutton Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, 22,Nº 3:400–407, 1951.
- [7] Robert H Shumway and David S Stoffer. Times series analysis and its applications with r examples. *Second Edition Springer Science Business Media LLC*, 2006.
- [8] James C Spall. A stochastic approximation technique for generating maximum likelihood parameter estimates. *The Johns Hopkins University Laurel, Maryland*, pages 1161–1167, 1987.
- [9] James C. Spall. Adaptive stochastic approximation by the simultaneous perturbation method. *IEEE Transactions on Automatic Control*, 45:1839–1853, 2000.
- [10] Jesús Fernández Villaverde and Juan F. Rubio Ramirez. Estimating dynamic equilibrium economies: Linear versus nonlinear likelihood. *Journal of applied econometrics*, 20:891–910, 2005.

- [11] Xun Zhu and James Spall. A modified second order spsa optimization algorithm for finite samples. applied physics laboratory. *International Journal of adaptive control and signal processing*, 16:397–409, 2000.