



TECNOLOGICO
DE MONTERREY®



TECNOLOGICO
DE MONTERREY

Biblioteca
Campus Ciudad de México

Instituto Tecnológico y de Estudios
Superiores de Monterrey
Campus Ciudad de México

*Análisis de componentes principales y
prueba de especificación no paramétrica
de las tasas de interés en México*

TESIS QUE PARA RECIBIR EL GRADO DE
DOCTOR EN CIENCIAS FINANCIERAS
PRESENTA

Carlos Armando Martínez Reyes

Director de tesis:
Dr. José Antonio Núñez Mora

México D.F., 26 Octubre 2011

Resumen

En esta investigación se describen los diferentes factores que se encuentran presentes en la estructura de plazos de las tasas de interés en México. Tomando como período de referencia de julio de 2002 a diciembre de 2009 y mediante un análisis de componentes principales aplicado a los cambios diarios de los CETES, encontramos que para analizar la dinámica de la curva de rendimiento, es posible simplificar el estudio a tan solo dos factores que se hallan latentes en los datos: el nivel y la pendiente. El factor preponderante el cual explica un poco más del 68% de los cambios se halla asociado al nivel de la curva, y está determinado principalmente por la tasa de interés de menor plazo y por choques estructurales que desplazan paralelamente a la curva. Por su parte la pendiente que explica el 27,6%, usualmente se calcula como la diferencia entre el rendimiento de mayor plazo y el de menor vencimiento. Los cambios en la pendiente, implican que factores macroeconómicos, tales como el nivel de precios de la economía, la actividad económica y las tasas de interés a corto plazo (instrumentos con vencimientos menores a un año) están detrás de ese movimiento.

Una vez identificados el número de factores necesarios para describir los cambios en la curva de rendimiento, proponemos una prueba para la especificación de modelos paramétricos de dos factores. Se presentan tres pruebas diferentes. Las dos primeras se basan en comparar la función de densidad marginal obtenida a partir del modelo paramétrico de la estructura temporal y un estimador no paramétrico de la función de densidad marginal obtenida a partir de los datos. La última prueba se basa en la idea de comparar el estimador de la densidad kernel y el modelo paramétrico del kernel suavizado para evitar los efectos del sesgo. La ventaja de la última prueba con respecto a los dos primeras es que

es válido para datos de bajo-suavizado y datos de sobre-suavizado. La prueba se puede aplicar a muchos procesos de interés en finanzas. En particular, la aplicamos para determinar si la dinámica de la estructura temporal de los CETES se puede modelar a partir de los modelos de Brennan-Schwartz (1979) y Schaefer-Schwartz (1984), los resultados de la prueba nos dicen que ambos modelos continuos son rechazados al nivel del 5% con el fin de describir la dinámica de los CETES.

Índice general

Resumen	II
Lista de figuras	VI
Lista de tablas	VII
1. Introducción	1
2. Estructura temporal de las tasas de interés: conceptos básicos.	3
2.1. Teorías de la curva de rendimiento	4
2.2. Modelos Multifactoriales de la estructura temporal de la tasa de interés. . .	5
3. Revisión de conceptos sobre procesos de difusión.	10
3.1. Procesos de difusión	10
3.2. Métodos de estimación.	14
3.2.1. Estimación no paramétrica: métodos de suavizado.	15
4. Dinámica de la curva de rendimiento en México: un análisis de componentes principales.	22
4.1. Análisis de componentes principales	22
4.1.0.1. Análisis factorial vs análisis de componentes principales	22
4.1.0.2. Desarrollo del análisis de componentes principales	23
4.1.1. Revisión de la literatura	26
4.2. Metodología	30
4.2.1. Datos	30

4.2.2. Estadística descriptiva: niveles	32
4.2.3. Estadísticas descriptivas: diferencias	34
4.2.4. Condiciones para aplicar el análisis de componentes principales.	35
4.2.5. Aplicación del ACP	40
4.2.5.1. Configuración general	40
4.2.5.2. Resultados.	42
4.2.5.3. Número de componentes principales a retener.	44
4.2.5.4. Interpretación de los componentes principales.	46
5. Prueba de especificación no paramétrica para modelos de tiempo conti-	
nuo.	52
5.1. Introducción.	52
5.2. El modelo y la hipótesis nula.	54
5.3. La distribución límite de I_n bajo H_0 .	56
5.4. Prueba estadística y sus distribuciones asintóticas.	62
5.5. Extensión del estadístico de prueba de Yacine Aït-Sahalia.	63
5.6. Estadístico de prueba con corrección del sesgo.	64
5.7. Modelos paramétricos de dos factores.	65
5.7.1. Modelo de la estructura de plazos de Brennan-Schwartz y Schaefer-	
Schwartz.	65
5.7.1.1. Modelo de la estructura de plazos de Brennan - Schwartz .	65
5.7.1.2. Modelo de la estructura de plazos de Schaefer-Schwartz . .	67
5.7.2. Pruebas empíricas y análisis de especificación de la estructura de	
plazos de los modelos de Brennan-Schwartz y Schaefer-Schwartz . .	68
6. Conclusiones	71

Índice de figuras

4.1. Observaciones diarias CETES: julio 2002 a diciembre 2009	31
4.2. Rendimientos de los CETES: julio 2002 a diciembre 2009	32
4.3. Volatilidad anualizada de los CETES en función de los distintos vencimientos.	33
4.4. Gráfico de sedimentación.	46
4.5. Primer componente principal.	47
4.6. Comportamiento del 1er componente y el promedios total de las series de rendimientos.	48
4.7. Segunda componente principal.	49
4.8. Comportamiento del 2do componente y el spread rendimiento a 10 años menos rendimiento a 1 día.	49
4.9. Tercer componente principal.	50
4.10. Vectores característicos de los tres primeros componentes principales.	51

Índice de tablas

2.1. Modelos de varios factores de la estructura temporal.	6
4.1. Diferencias entre el análisis de componentes principales y el análisis de factores.	23
4.2. Estudios sobre el ACP y el análisis factorial en mercados extranjeros.	27
4.3. Estadística descriptiva: niveles	33
4.4. Estadística descriptiva: diferencias.	34
4.5. Matriz de correlaciones: cambios diarios	35
4.6. Matriz de correlación: niveles	36
4.7. KMO y prueba de Bartlett.	37
4.8. Pruebas de estacionariedad para los niveles y primeras diferencias de los CETES.	39
4.9. Valores críticos para las pruebas de estacionariedad.	39
4.10. Valores propios y porcentaje de varianza explicada por los cambios de las tasas de interés.	42
5.1. Valores de los parámetros estimados en el modelo de Brennan y Schwartz	69
5.2. Valores de los parámetros estimados en el modelo de Schaefer y Schwartz .	69
5.3. Pruebas de especificación paramétrica de modelos de tiempo continuo de dos factores.	70

Introducción

LA estructura temporal de las tasas de interés ha recibido una atención considerable durante las últimas décadas tanto por profesionales como por investigadores, esto se debe a las numerosas aplicaciones que tiene. Desde el punto de vista de la Economía Financiera, permite valorar múltiples activos y diseñar estrategias de inversión o de cobertura. En la Teoría Económica, es utilizada en el estudio de temas tales como la formación de expectativas, las relaciones entre las tasas a corto y largo plazo, la transmisión de la política monetaria a las variables macroeconómicas relevantes. Un elemento clave de toda estructura de plazos es la existencia de un modelo que generalmente viene representado por una ecuación diferencial estocástica cuya solución es un proceso de difusión que determina la dinámica que sigue la tasa de interés instantánea. Existen varias maneras de estimar un proceso de difusión, cada una con sus fortalezas y debilidades. En años anteriores las investigaciones se han orientado hacia la especificación parsimoniosa de la función de tendencia y difusión, con la finalidad de que la ecuación diferencial puede ser resuelta analíticamente.

Otra posibilidad es emplear una parametrización flexible como la del enfoque de Aït-Sahalia (1996) en la cual se construye un estimador y una prueba de especificación basada en la minimización de la distancia entre un estimador de la densidad marginal implícita en el modelo y un estimador no paramétrico de densidad kernel. A pesar de que este enfoque sufre de pérdida de información debido al uso de una densidad marginal, la sugerencia de Aït-Sahalia (1996) sobre la especificación de una densidad de transición, conduce a considerables dificultades de cálculo en la aplicación de la prueba.

El resto del trabajo se organiza de la siguiente manera. En el capítulo 2 definimos la estructura temporal de las tasas de interés e introducimos las teorías de la curva de rendimiento y los modelos multifactoriales que describen su dinámica, en el capítulo 3 damos los antecedentes básicos de los procesos de difusión y una breve introducción sobre estimación no paramétrica. En el capítulo 4 de este trabajo se utiliza la técnica estadística de análisis de componentes principales para identificar los factores que conducen la estructura temporal de los bonos cupón cero (CETES) del Gobierno de México en el período de julio de 2002 a diciembre de 2009. El objetivo de este ejercicio es hallar e identificar las principales fuentes de variación de la estructura temporal de los CETES en el periodo señalado. En el capítulo 5, se extiende la prueba de Aït-Sahalia (1996) de modelos de tiempo continuo de un factor a modelos de la estructura de plazos de dos factores que tienen funciones de densidad marginal estacionaria. Estudiamos, cómo la prueba de Sahalia (1996) y nuestra prueba ampliada dependen de la cantidad de suavizado que poseen los datos. Creemos que el sesgo introducido en la estimación de la densidad Kernel tiene una influencia significativa en estas pruebas. En consecuencia, con el fin de eliminar la influencia del sesgo, se propone una prueba estadística de sesgo corregido sobre la base de comparar el estimador de la densidad kernel y la función de densidad con kernel suavizado generada por el modelo paramétrico. Posteriormente, aplicamos la prueba ampliada y la prueba de sesgo corregido para probar las especificaciones a dos casos especiales: el modelo de Brennan-Schwartz (1979) y el modelo de Schaefer-Schwartz (1984), que pertenecen a la clase afín de los modelos de estructura de plazos desarrollados por Duffie y Kan (1996).

Estructura temporal de las tasas de interés: conceptos básicos.

LA relación entre las tasas de interés de bonos de la misma calidad crediticia y el tiempo que falta para su vencimiento es conocida como Estructura Temporal de las Tasas de Interés (ETTI) o estructura a plazos, ésta juega un papel fundamental en los mercados financieros ya que contiene información valiosa para el diseño de políticas monetarias. Así mismo, refleja de una manera clara y sencilla las expectativas de los agentes acerca de la trayectoria futura de las tasas de interés, la prima por riesgo y la liquidez de los instrumentos que conforman un mercado de capitales.

La curva de rendimiento o *yield curve* es la representación gráfica de la estructura temporal de las tasas de interés en diferentes momentos del tiempo. La importancia del estudio de esta curva radica en que la ETTI de los rendimientos contiene información sobre las decisiones de ahorro e inversión de los agentes económicos en el tiempo y, por consiguiente, de la demanda que es un elemento que ayuda a determinar la inflación y otros indicadores. La curva de rendimiento permite estimar la estructura temporal de las tasas de interés para activos financieros homogéneos en cuanto a sus características intrínsecas tales como riesgo, régimen fiscal, emisor, mercado y permite observar la gama de rendimientos ofrecidos por dichos activos teóricamente idénticos en todo, excepto en su plazo de vencimiento.

2.1. Teorías de la curva de rendimiento

La teoría de las expectativas y la teoría de la segmentación del mercado constituyen las principales explicaciones de la forma de la curva de rendimiento. A su vez, dependiendo de los supuestos subyacentes, la teoría de las expectativas puede dividirse en la teoría de las expectativas puras, teoría de la preferencia por liquidez y la teoría del hábitat preferido.

1. Teoría de las expectativas puras:

Según esta teoría, la forma de la curva de rendimiento se determina por las expectativas de los agentes acerca de la trayectoria futura de las tasas de interés, las cuales, debido al proceso de arbitraje entre las diferentes opciones de inversión, constituyen el mejor estimador disponible del mercado (Fisher, 1896; Hicks, 1939). Dicha teoría sugiere que la tasa de interés a largo plazo es el promedio aritmético (geométrico) de las tasas a corto plazo vigentes y esperadas.

2. Teoría de la preferencia por liquidez (Hicks, 1939):

En un ambiente de incertidumbre y aversión al riesgo, los agentes prefieren los bonos de menor madurez por tener una mayor liquidez. Por lo tanto, los inversionistas aceptarán tener en sus portafolios títulos a largo plazo, únicamente si dichos activos ofrecen una compensación (prima de liquidez) representada en una tasa de interés mayor, la prima de liquidez ofrecida siempre será una función creciente del plazo al vencimiento. De esta manera, la teoría de la liquidez plantea que la curva de rendimiento normalmente exhibirá una pendiente positiva, incluso si el mercado no espera cambios en las tasas de interés.

3. Teoría de la segmentación del mercado (Cultberson, 1957):

Al igual que la teoría del hábitat, establece que los inversionistas tienen lugares predilectos en los cuales ajustan perfectamente los vencimientos de los pasivos con los de sus activos. Sin embargo, sostiene que prestamistas y prestatarios se limitan a ciertos segmentos de la curva de acuerdo con la regulación, el costo de la información,

entre otros, siendo indiferentes a las primas de riesgo ofrecidas por diferentes activos financieros. Las tasas de cada punto de la curva de rendimiento se determinan por las condiciones de oferta y demanda de cada segmento.

4. Teoría del hábitat preferido (Modigliani y Sutch, 1996):

Los agentes con aversión al riesgo eliminan todo riesgo sistemático al situarse en su hábitat preferido que es aquel tramo de la curva de rendimiento en el que la vida media de sus activos coincide con la de sus pasivos (inmunización de portafolios). El rendimiento para cada plazo de inversión es una función de la oferta y la demanda de recursos para dicho horizonte temporal. Debido a la segmentación del mercado, en aquellos vencimientos donde la demanda de fondos es insuficiente, los emisores de bonos deberán ofrecer tasas de interés mayores que incorporen una prima de riesgo para lograr así que los agentes abandonen su hábitat preferido.

2.2. Modelos Multifactoriales de la estructura temporal de la tasa de interés.

Una cuestión que se plantea, a la hora de elegir modelos con más de un factor, es cuántos factores deberían ser tomados en cuenta desde un punto de vista práctico. Estudios históricos de la curva de rendimientos, basados en el análisis de componentes principales, sugieren que una variable explica entre el 80% y el 90% del total de la varianza mientras que dos variables explican entre el 95% y el 99% de la misma, Rebonato (1996). Estos valores pueden presentar una cierta variación dependiendo del mercado financiero que se analice. Por tanto, la elección del número de factores supone un compromiso entre la implementación numérica eficiente del modelo, y su capacidad para representar la correlación de forma realista y ajustarse a los datos del mercado satisfactoriamente.

A continuación presentamos los modelos multifactoriales más conocidos en la literatura.

Richard (1978): supone que la dinámica de la estructura de plazos de las tasas de interés viene determinada por dos variables: la tasa de interés real instantánea, R , y la tasa

Autor	Modelo
Richard (1978)	$dR = \beta_R(m_R - R)dt + \rho_R\sqrt{R}dw$ $d\pi = \beta_\pi(m_\pi - \pi)dt + \rho_\pi\sqrt{\pi}dz_\phi$
Brennan y Schwartz(1982)	$dr = (a_r + b_r(l - r))dt + \rho_r dw$ $dl = (a_l + b_l r + c_l l)dt + l\sigma_l dz_l$
Boyle (1980)	$dR = \beta_R(m_R - R)dt + \rho_R dw$ $d\pi = \beta_\pi(m_\pi - \pi)dt + \rho_\pi dz_\pi$
Langtieg (1980)	$dx_i = \beta_i(m_i - r)dt + \rho_i dw_i, r = \sum_{i=1}^n x_i$
Shaefer y Schwartz (1984)	$dS = \beta_s(m_s - s)dt + \rho_s dw$ $dl = \beta(s, l, t)dt + \rho_l \sqrt{l} dz_l$
Fong y Vasiceck (1991)	$dr = \beta(\bar{r} - r)dt + \sqrt{v}dw$ $dv = \gamma(\bar{v} - v)dt + \epsilon\sqrt{v}dz_v$
Longstaff y Schwartz (1992)	$dx = (a_x - b_1 x)dt + c_x \sqrt{x}dw$ $dy = (a_y - b_y y)dt + c_y \sqrt{y}dz_y$
Duffe y Kan (1996)	$dx_i = (a_i + b_i x_i)dt + \sqrt{c_i + d_i x_i}dw$
Chen (1996)	$dr = \beta_r(\theta - r)dt + \sqrt{v}dw$ $d\theta = \beta_\theta(\bar{\theta} - \theta)dt + \rho_\theta \sqrt{\theta} dz_\theta$ $dv = \beta_\nu(\bar{v} - v)dt + \rho_\nu \sqrt{v} dz_\nu$
Chen y Scott (1996)	$r = y_1 + y_2$ $dy_i = \beta(m_i - y_i)dt + \rho_i \sqrt{y_i} dw_i, i = 1, 2$

Tabla 2.1: Modelos de varios factores de la estructura temporal.

de inflación instantánea esperada, π . La dinámica en el tiempo de estas dos variables consiste en que ambas variables son procesos de tipo raíz cuadrada independientes. En cuanto a los precios del riesgo de mercado, supone que tienen la misma estructura que el utilizado en el modelo unifactorial de Cox, Ingersoll, y Ross (1985). La ventaja fundamental es que proporciona una solución analítica, sin embargo, R y π son difíciles de observar y de modelar. Este modelo también es planteado por Cox, Ingersoll, y Ross (1985) de forma teórica, sin asignar una interpretación económica a las variables.

Brennan y Schwartz (1982): consideran que el comportamiento de la curva de rendimientos se puede explicar a partir de dos variables a priori desconocidas. En particular, si las tasas de interés a corto y largo plazo se pueden expresar invirtiendo su dependencia implícita como funciones diferenciables, entonces el análisis se puede realizar considerando a estas dos como las variables de estado. Para realizar un análisis cuantitativo del modelo, Brennan y Schwartz (1980b), y Brennan y Schwartz

(1982) suponen que la dinámica de las tasas de interés tiene un comportamiento localmente lognormal. Además, consideran que la tasa de interés a corto plazo tiende hacia el largo plazo con una velocidad b_r . Respecto al premio al riesgo de mercado de la tasa de interés a corto plazo, suponen que es constante, y en cuanto al premio al riesgo de mercado asociado a la tasa de interés a largo plazo, no aparece ningún parámetro en el modelo, ya que consideran que la tasa de interés a largo plazo es el rendimiento de un bono perpetuo, el cual es un título negociable. Los principales inconvenientes de este modelo son los siguientes. En primer lugar su inestabilidad, ya que puede explotar con probabilidad positiva en tiempo finito, es decir, las tasas de interés a corto y largo plazo pueden llegar a infinito con probabilidad positiva, Rebonato (1996). En segundo lugar, no proporciona una solución analítica sino que son necesarios métodos numéricos para obtener una solución aproximada.

Boyle (1980): plantea un modelo de la estructura temporal de las tasas de interés basado en las mismas variables de estado que Richard (1978), pero a diferencia de éste, supone que ambas variables siguen un proceso con volatilidad constante y que los precios del riesgo de mercado son constantes, hipótesis utilizadas por Vasicek (1977) en su modelo unifactorial. De esta forma, obtiene también una solución cerrada para los precios de los bonos cupón cero.

Langtieg (1980): extendió el modelo de Vasicek (1977) suponiendo que la tasa de interés se puede descomponer como la suma de n factores que siguen un proceso Ornstein-Uhlenbeck. Langtieg proporciona solución exacta para los precios de los bonos cupón cero y para las opciones sobre estos bonos.

Schaefer y Schwartz (1984): presentan también un modelo de la estructura temporal de dos factores, pero lo expresan en términos de la tasa de interés a largo plazo, l , y el spread o diferencia entre e la tasa de interés a corto y a largo plazo, s . La elección de estas variables se basa en la evidencia empírica de ortogonalidad entre dl y ds , lo cual permite obtener una solución aproximada para el precio de los bonos cupón cero a partir de la solución exacta de la ecuación en derivadas parciales modificada.

Fong y Vasiceck (1991), Fong y Vasiceck (1992a) y Fong y Vasiceck (1992b): en una serie de trabajos por un lado, y Longstaff y Schwartz (1992), por otro, desarrollan los modelos utilizando la tasa de interés a corto plazo y la varianza de los cambios de la tasa de interés, pero con una dinámica y un planteamiento diferente. En los trabajos de Fong y Vasiceck (1991), Fong y Vasiceck (1992a) y Fong y Vasiceck (1992b), la obtención del precio de los bonos cupón cero es laboriosa y requiere el uso de Algebra Compleja. Selby y Strickland (1995) proporcionan una aproximación eficiente por series.

Longstaff y Schwartz (1992): desarrollan un modelo de equilibrio general de la economía, a partir del cual se deduce uno de la estructura de plazos de dos factores. Parten de un inversionista representativo con función de utilidad logarítmica, y que tiene la posibilidad de elegir entre invertir o consumir el único bien disponible en la economía. Longstaff y Schwartz (1992) no proporcionan ninguna interpretación intuitiva para los factores que utilizan, no obstante, demuestran que se pueden relacionar con otras variables financieras observables como son la tasa de interés instantánea, r , y su varianza, σ . Este modelo de equilibrio da lugar a uno afín, y es posible obtener una solución exacta para los bonos cupón cero y las opciones europeas sobre estos bonos. Sin embargo, su implementación práctica es muy compleja debido a las dificultades que presenta la estimación de los numerosos parámetros.

Duffie y Kan (1996): introdujeron una nueva clase de modelos de la estructura temporal en los cuales la tendencia y la volatilidad de los procesos estocásticos de las variables de estado son afines. Suponen que los factores son los rendimientos de varios bonos cupón cero con diferentes vencimientos. Cada uno de los rendimientos se define como un proceso de Markov, es observable, y sus incrementos pueden tener una correlación arbitrariamente especificada con otros rendimientos. Los precios de los bonos cupón cero se obtienen resolviendo un sistema de ecuaciones diferenciales ordinarias. Esta clase de modelos son muy interesantes desde un punto de vista analítico.

Chen (1996): considera tres variables de estado que son; la tasa de interés instantánea r ,

el valor medio esperado al que tiende la tasa de interés instantánea θ , y su volatilidad ν . En cuanto a la tasa de interés instantánea, supone que es un proceso con volatilidad constante, pero los parámetros del modelo son a su vez variables estocásticas de tipo raíz cuadrada. Cuando se supone independencia de las variables, se conoce su solución analítica.

Chen y Scott (1996): proponen un modelo en el que la tasa de interés nominal se obtiene como la suma de dos factores independientes que siguen procesos de tipo raíz cuadrada. Estos autores suponen diferentes significados económicos para estas variables basándose en el trabajo de Cox, Ingersoll, y Ross (1985). Este modelo presenta la ventaja de que proporciona soluciones analíticas para varios activos derivados de la tasa de interés.

Revisión de conceptos sobre procesos de difusión.

Los procesos de Markov constituyen la clase mas importante de procesos estocásticos que son útiles para modelar fenómenos con cierto tipo de dependencia, que aparece con frecuencia en diversas aplicaciones. Hay dos temas que discutiremos en esta capítulo. Comenzamos con una presentación de algunos resultados esenciales de la teoría de difusiones, especialmente los teoremas de Feynman-Kac y de Girsanov. En segundo lugar consideramos los métodos de estimación no paramétrica, en particular los métodos de suavizado.

3.1. Procesos de difusión

Definición 1. *Un proceso estocástico de tiempo continuo $X = \{X_t, t \geq 0\}$ es llamado proceso de Markov si satisface la propiedad de Markov, i.e.*

$$P(X(t_{n+1}) \in B \mid X(t_1) = x_1, X(t_2) = x_2, \dots, X(t_n) = x_n) = P(X(t_{n+1}) \in B \mid X(t_n) = x_n) \quad (3.1)$$

Para todo subconjunto Boreliano de \mathbb{R} , y todos los instantes de tiempo $0 < t_1 < t_2 < \dots < t_n < t_{n+1}$ y todos los estados $x_1, x_2, \dots, x_n \in \mathbb{R}$ para los cuales las probabilidades condicionales están definidas.

En la práctica, observamos solamente un resultado del proceso, una trayectoria, de una gran cantidad de posibles trayectorias. Una trayectoria es una función del tiempo $x_t = x(t); 0 \leq t \leq T$. Para hacer cálculos de un futuro incierto, se necesita conocer las

propiedades colectivas de todas las trayectorias lo que se logra con las distribuciones de probabilidad de la variable aleatoria $X(t)$.

Sea X un proceso de Markov, escribimos sus probabilidades de transición como

$$\mathbf{P}(s, x, ; t, A) = P(X(t) \in A \mid X(s) = x), \quad 0 \leq s < t \quad (3.2)$$

Si la distribución de probabilidad P es discreta, las probabilidades de transición están determinadas de manera única por la matriz de transición con componentes

$$\mathbf{P}(s, i; t, j) = P(X(t) = x_j \mid X(s) = x_i) \quad (3.3)$$

que dependen también de s y t , mientras que en el caso continuo

$$\mathbf{P}(s, x; t, A) = \int_A \mathbf{p}(s, x; t, y) dy \quad (3.4)$$

para todo $A \in \mathcal{B}$, donde la densidad $p(s, x; t, \cdot)$ es llamada la *densidad de transición*.

Un proceso de Markov se dice que es *homogéneo* si todas sus probabilidades de transición dependen solamente de la diferencia de tiempo $t - s$ más que de sólo los valores específicos de s y t .

El proceso de Wiener y los procesos Ornstein-Uhlenbeck con parámetro $\gamma = 1$, son procesos de Markov homogéneos y sus densidades de transición son respectivamente:

$$\mathbf{p}(s, x; t, y) = \frac{1}{\sqrt{2\pi(t-s)}} \exp\left(-\frac{(y-x)^2}{2(t-s)}\right) \quad (3.5)$$

$$\mathbf{p}(s, x; t, y) = \frac{1}{\sqrt{2\pi(1-e^{-2(t-s)})}} \exp\left(-\frac{(y-xe^{-(t-s)})^2}{2(1-e^{-2(t-s)})}\right) \quad (3.6)$$

Como $p(s, x; t, y) = p(0, x; t-s, y)$ usualmente omitimos la primera variable y simplemente escribimos $\mathbf{P}(x; t-s, y)$, con $\mathbf{P}(x; t-s, B) = \mathbf{P}(0, x; t-s, B)$ para tales probabilidades de transición.

En conclusión tenemos la siguiente definición:

Definición 2. Sea $p(s, x, t, A)$ una función no negativa definida para $0 \leq s < t < \infty$, $x \in \mathbb{R}$, $A \in \mathcal{B}$ tal que:

- (i) $p(s, x, t, A)$ es Borel medible en x , para s, t, A fijos;
- (ii) $p(s, x, t, A)$ es una medida de probabilidad en A , para s, x, t fijos;
- (iii) p satisface la ecuación de Chapman-Kolmogorov

$$p(s, x, t, A) = \int_{\mathbb{R}} p(s, x, \lambda, dy) p(\lambda, y, t, A) \text{ para cualquier } s < \lambda < t \quad (3.7)$$

Entonces llamamos a p una *función de transición de Markov*, una *función de probabilidad de transición*, o simplemente *probabilidad de transición*.

Luego, la función de **densidad de transición** del proceso de difusión $\{X_t\}_{t \geq 0}$ es la función $p : \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}_+ \times \mathcal{B}(\mathbb{R}) \rightarrow \mathbb{R}_+$ para la cual

$$\mathbb{P}[X_t \in B \mid X_0 = x] \equiv \mathbb{P}_x[X_t \in B] = \int_B p(s, x, t, A) dt \quad (3.8)$$

para cualquier subconjunto $B \subseteq \mathbb{R}$

La existencia de tal función es garantizada para todos los procesos que consideraremos a lo largo del trabajo. Note que $p(s, x, t, A)$ es justo la función de densidad de probabilidad para la posición de X_t dado que $X_0 = x$. Se pueden considerar procesos más generales, pero solamente estamos interesados en difusión que son *homogéneas en el tiempo*, es decir, tales que

$$\mathbb{P}[X_{t+s} \in A \mid X_s = x] = \mathbb{P}[X_t \in A \mid X_0 = x] \quad \forall s \geq 0, \forall x \in \mathbb{R} \quad (3.9)$$

La propiedad de Markov nos dice que para $s < t$

$$\mathbb{P}_x[X_t \in A \mid \{X_r\}_{0 \leq r \leq s}] = \mathbb{P}_x[X_t \in A \mid X_s] \quad (3.10)$$

y en consecuencia tenemos que la ecuación de Chapman-Kolmogorov cumplen lo siguiente:

Lema 1 (Ecuaciones de Chapman-Kolmogorov). *Para $s < t$, $(x, z) \in \mathbb{R} \times \mathbb{R}$*

$$p(t, x, z, A) = \int_{\mathbb{R}} p(t - s, x, y, A)p(s, y, z, A)dy \quad (3.11)$$

Los procesos de difusión, los cuales definiremos en el caso de una dimensión, son una rica y útil clase de procesos de Markov.

Definición 3. *Un proceso de Markov con densidad de transición $\mathbf{P}(s, x; t, y)$ es llamado un **proceso de difusión** si los siguientes tres límites existen para toda $\epsilon > 0, s \geq 0$ y $x \in \mathbb{R}$.*

$$\lim_{t \downarrow s} \frac{1}{t - s} \int_{|y-x| > \epsilon} \mathbf{P}(s, x; t, y)dy = 0 \quad (3.12)$$

$$\lim_{t \downarrow s} \frac{1}{t - s} \int_{|y-x| < \epsilon} (y - x)\mathbf{P}(s, x; t, y)dy = \mu(s, x) \quad (3.13)$$

$$\lim_{t \downarrow s} \frac{1}{t - s} \int_{|y-x| < \epsilon} (y - x)^2\mathbf{P}(s, x; t, y)dy = \sigma^2(s, x) \quad (3.14)$$

donde μ y σ son funciones bien definidas.

La condición (3.12) previene que el proceso de difusión tenga saltos instantáneos, la cantidad μ es llamada drift y σ es el coeficiente de difusión en el tiempo s y la posición x . (3.13) implica que

$$\mu(s, x) = \lim_{t \downarrow s} \frac{1}{t - s} \mathbb{E}(X(t) - X(s) \mid X(s) = x) \quad (3.15)$$

así que el drift $\mu(s, x)$ es la tasa instantánea de cambio en la media del proceso dado que $X(s) = x$. Similarmente, se sigue de (3.14) que el coeficiente de difusión al cuadrado

$$\sigma^2(s, x) = \lim_{t \downarrow s} \frac{1}{t - s} \mathbb{E}((X(t) - X(s))^2 \mid X(s) = x) \quad (3.16)$$

denota la tasa instantánea de cambio de las fluctuaciones al cuadrado del proceso dado

que $X(s) = x$.

Cuando el drift o tendencia μ y el coeficiente de difusión σ de un proceso de difusión son funciones suficientemente suaves, la densidad de transición $\mathbf{P}(s, x; t, y)$ también satisface una ecuación diferencial parcial, llamada *Ecuación de Kolmogorov adelantada*

$$\frac{\partial p}{\partial t} + \frac{\partial}{\partial y} \{ \mu(t, y)p \} - \frac{1}{2} \frac{\partial^2}{\partial y^2} \{ \sigma^2(t, y)p \} = 0, \quad (s, x) \text{ fijo} \quad (3.17)$$

y la *Ecuación de Kolmogorov retrasada*

$$\frac{\partial p}{\partial s} + \mu(s, x) \frac{\partial p}{\partial x} + \frac{1}{2} \sigma^2(s, x) \frac{\partial^2 p}{\partial x^2} = 0, \quad (t, y) \text{ fijo} \quad (3.18)$$

La ecuación (3.17) es comunmente llamada la ecuación de *Fokker-Planck*.

El ejemplo más simple de un procesos de difusión es el proceso conocido como movimiento Browniano, la manera más sencilla de pensar en un movimiento Browniano es un límite re-escalado de una caminata aleatoria. Sea, $\{Z_i\}_{i \in \mathbb{N}}$ variables aleatorias independientes idénticamente distribuidas con $\mathbb{P}[Z_i = 1] = \frac{1}{2} = \mathbb{P}[Z_i = -1]$, se define $S_n = \sum_{i=1}^n Z_i$. Si tomamos $S_0 = 0$, $\{S_n\}_{n \geq 0}$ es el proceso conocido como caminata aleatoria simple sobre \mathbb{Z} .

3.2. Métodos de estimación.

En el campo de la estructura de plazos se han aplicado en primer lugar las técnicas paramétricas, y posteriormente, las no paramétricas. La diferencia fundamental que existe entre ambas reside en que en las primeras, las funciones a estimar tienen una expresión conocida y dependen de un conjunto de parámetros que son los que se estiman. Dentro de estas técnicas el *método generalizado de momentos* es uno de los más empleados en la literatura para estimar los parámetros del proceso estocástico. En cuanto a las no paramétricas, se basan en la estimación de las diferentes funciones que aparecen en el modelo,

sin determinar a priori ninguna relación funcional. Dentro de estas técnicas está el *método del kernel*. La utilización de las técnicas no paramétricas presenta ciertas ventajas frente a las paramétricas, pero también ciertos inconvenientes. Las primeras son versátiles y flexibles ya que no es necesario especificar formas funcionales a priori, y mejoran el ajuste. Sin embargo, su funcionamiento no es muy adecuado, si existen muchos regresores o si el número de observaciones del que disponemos no es elevado. Además, tienden a sobreajustar las funciones. Para valorar si una estimación paramétrica es adecuada o no, existe una gran variedad de estadísticos que nos proporcionan dicha información. En el caso de la estimación no paramétrica, resulta muy útil construir las bandas de variabilidad de las funciones estimadas y, para ello, se aplica el algoritmo de bootstrap por bloques, que permite conservar la estructura de correlación de los datos.

3.2.1. Estimación no paramétrica: métodos de suavizado.

La mayoría de los modelos de estructura temporal de las tasas de interés, estudiados en el capítulo 3, se basan en una serie de hipótesis sobre el posible comportamiento de la tendencia y de la volatilidad del proceso estocástico que recoge la dinámica de las tasas de interés. Sin embargo, no existe evidencia empírica clara sobre qué comportamiento es el más adecuado; de hecho, no se ha podido demostrar todavía que un determinado tipo de proceso estocástico sea el mejor para modelar el comportamiento de las tasas de interés. En los diferentes modelos paramétricos que se conocen, se ha intentado capturar diferentes propiedades particulares de la dinámica de las tasas de interés.

Esto ha llevado a utilizar técnicas de estimación no paramétrica, para capturar una gran variedad de no linealidades sin necesidad de especificar relación paramétrica alguna. A diferencia de los métodos paramétricos, los no paramétricos requieren menos hipótesis y restricciones, pero también presentan inconvenientes. Por ejemplo, para su correcto funcionamiento se requiere un gran número de datos y, en general, tienden a sobreajustar las funciones.

Existen numerosas técnicas de estimación no paramétricas, pero quizás las más utilizadas sean los *métodos de suavizado*, en los cuales los errores de observación se reducen ponderando los datos de diferentes formas.

La idea fundamental de estos métodos consiste en suavizar un conjunto de datos $\{(X_i, Y_i)\}_{i=1}^n$, de la curva $m(x)$, en la siguiente relación de regresión:

$$Y_i = m(X_i) + \varepsilon_i \quad i = 1, \dots, n \quad (3.19)$$

donde $\{\varepsilon_i\}_{i=1}^n$ son variables aleatorias de media cero y varianza σ^2 . Así, si $\{(X_i, Y_i)\}_{i=1}^n$ son variables aleatorias independientes e igualmente distribuidas, la curva de regresión se define como

$$m(x) = \mathbb{E}[Y|X = x] \quad (3.20)$$

con $\mathbb{E}|Y| < \infty$. La función $m(x)$ no tiene ninguna restricción en cuanto a su comportamiento, no tiene porque poder representarse como una combinación de parámetros y puede ser no lineal, tan solo requiere que tenga p derivadas continuas (es decir, que sea suave) con respecto a los errores ε_i .

Si existe la función de densidad conjunta $f(x, y)$, entonces $m(x)$ se puede obtener como

$$m(x) = \int \frac{f(x, y)y}{f(x)} dy \quad (3.21)$$

donde $f(x) = \int f(x, y)dy$ denota la función de densidad marginal de X .

El objetivo de la estimación no paramétrica de $m(x)$ en (3.19) es aproximar los datos $Y_i, i = 1, 2, \dots, n$, lo máximo posible a partir de una muestra lo suficientemente grande. La estimación de $m(x)$ se obtiene a partir de una clase general de estimadores lineales no paramétricos; la media ponderada de Y_i .

$$\hat{m}(x) = \sum_{i=1}^n W_{n,i}(x)Y_i \quad (3.22)$$

donde $W_{n,i} = W_n(X_i, x)$ representa el peso asignado a la observación i -ésima, que depende de la distancia del punto X_i al punto x .

Dentro de los métodos de suavizado podemos distinguir diferentes tipos, como por ejemplo, el **método del kernel**, el **desarrollo de series ortogonales** y el **método del k -vecino más próximo**. A continuación, vemos en qué consiste el método del kernel, y cómo autores como Stanton (1997) y Jiang (1998b) lo aplican a la estimación de la estructura de plazos de las tasas de interés.

El método del kernel es una representación muy sencilla de la sucesión de pesos $\{W_{n,i}(x)\}_{i=1}^n$. Este método describe la forma de la función peso mediante una función de densidad con un parámetro de escala que ajusta el tamaño y la forma de los pesos cerca de x . Esta función recibe habitualmente el nombre de kernel $K(x)$, es continua, acotada, real y simétrica. Además verifica

$$K(x) \geq 0 \tag{3.23}$$

$$\int K(u)du = 1 \tag{3.24}$$

A pesar de que la función $K(x)$ es una función de densidad de probabilidad, no desempeña un papel probabilístico, sino que únicamente se utiliza para calcular la media ponderada, y no implica que la distribución de la variable X tenga la distribución de probabilidad de $K(x)$.

La sucesión de pesos para el estimador kernel se define como

$$W_{n,i} = \frac{K_{h_n}(x - X_i)}{\hat{f}_{h_n}(x)} \tag{3.25}$$

con

$$K_{h_n}(u) = h_n^{-1} K\left(\frac{u}{h_n}\right) \tag{3.26}$$

es la función kernel con un factor de escala h_n . La función $\hat{f}_h(\cdot)$ es el estimador de la densidad kernel de Rosenblatt-Parzen de X . La forma (3.25) de los pesos fue propuesta por Nadaraya (1964), y Watson y Schwartz (1964) y recibe el nombre de *estimador de Nadaraya Watson*

$$\hat{m}(x) = \frac{n^{-1} \sum_{i=1}^n K_h(x - X_i) Y_i}{\sum_{i=1}^n K_h(x - X_i)} \quad (3.27)$$

La forma del kernel viene determinada por la función K ; sin embargo, el tamaño de los pesos viene parametrizado por h_n , que es el ancho de banda. Existe una gran variedad de funciones kernel y las más utilizadas, Härdle (1989), son la de Epanechnikov¹

$$K(u) = 0.75(1 - u^2)I(|u| \leq 1) \quad (3.28)$$

y la Gaussiana

$$K(u) = \frac{\exp(-u^2/2)}{\sqrt{2\pi}} \quad (3.29)$$

El parámetro de suavizado o anchura de banda, h_n , es de gran importancia en la estimación no paramétrica. La selección de este parámetro está relacionada con la interpretación que se realice del suavizado. Así, si deseamos presentar unos resultados o sugerir un modelo paramétrico, entonces una curva muy suave con un parámetro que puede ser elegido subjetivamente sería adecuado. Sin embargo, si lo que se desea es estimar una curva de regresión, entonces resulta adecuado que la curva no sea muy suave.

$$CV(h) = n^{-1} \sum_{i=1}^n [Y_j - \hat{m}_{h,j}(X_j)]^2 \omega(X_j) \quad (3.30)$$

con $\omega(X_j)$ una función peso no negativa que reduce los efectos frontera, Härdle (1989). El estimador $\hat{m}_{h,j}$ para la observación j -ésima se obtiene eliminado la observación x_j ,

$$\hat{m}_{h,j}(x) = \sum_{i \neq j} \omega_{n,i}(X_j) Y_i \quad (3.31)$$

En ocasiones, resulta también interesante estimar momentos de mayor orden. Sea $g(Y)$ una función suave que verifica que $\mathbb{E}|g(Y)| < \infty$ y sea $M(x) = \mathbb{E}(g(Y)|X = x)$ la media

¹En la fórmula (2.29) $I(|u| \leq 1)$ representa el valor de 1 si $|u| \leq 1$ y 0 en caso contrario.

condicional de la función $g(Y)$ dado $X = x$. Entonces un estimador de $M(x)$ se puede obtener de forma análoga a (3.22) que viene determinado por

$$\hat{M}(x) = \sum_{i=1}^n W_{n,i}(x)g(y_i) \quad (3.32)$$

donde los pesos $W_{n,i}$ están definidos en (3.25). Por ejemplo, si $g(Y) = Y^2$, $\hat{M}(x)$ proporciona la estimación no paramétrica del segundo momento condicionado centrado en el origen.

En la literatura relativa a la estructura temporal de las tasas de interés, existen varios autores que han presentado diferentes técnicas no paramétricas para estimar las funciones del proceso estocástico mediante el método del kernel.

Aït Sahalia(1996) propone una estimación semiparamétrica del proceso estocástico que siguen las tasas de interés. Considera un proceso con reversión a la media lineal y para su estimación utiliza el Método Generalizado de Mínimos Cuadrados. En cuanto a la volatilidad, la estima de forma no paramétrica utilizando la tendencia del proceso previamente calculada mediante técnicas paramétricas.

Basándonos en el resultado de Karatzas y Shreve (1991), por el cual la solución de la ecuación (3.1) es Markoviana siempre que las funciones $\alpha(r_t)$ y $\rho(r_t)$ estén acotadas en subconjuntos cerrados, la función de densidad de transición debe verificar la ecuación hacia adelante de Kolmogorov,

$$\frac{\partial}{\partial \Delta} p(\Delta, r_{t+\Delta}|r_t) = \frac{1}{2} \frac{\partial^2}{\partial r_{t+\Delta}^2} [\rho^2(r_{t+\Delta})p(\Delta, r_{t+\Delta}|r_t)] - \frac{\partial}{\partial r_{t+\Delta}} [\alpha(r_{t+\Delta})p_t(\Delta, r_{t+\Delta}|r_t)] \quad (3.33)$$

con $p(\Delta, r_{t+\Delta}|r_t)$ la correspondiente función densidad de transición, es decir, la función de densidad de $r_{t+\Delta}$ condicionada al valor de r_t .

Multiplicando a ambos lados de la igualdad (3.34) por la función de densidad de la tasao de interés, $\pi(r)$, e integrando con respecto a la tasa de interés, obtenemos que la

volatilidad del proceso verifica

$$\frac{d^2}{dr_{t+\Delta}^2}[\rho^2(r_{t+\Delta})\pi(r_{t+\Delta})] = 2\frac{d}{dr_{t+\Delta}}[\alpha(r_{t+\Delta})\pi(r_{t+\Delta})] \quad (3.34)$$

Integrando dos veces, y usando la condición de frontera $\pi(0) = 0$, llegamos a que

$$\rho^2(r_{t+\Delta}) = \frac{2}{\pi(r_{t+\Delta})} \int_0^{r_{t+\Delta}} \alpha(u)\pi(u)du \quad (3.35)$$

Para obtener la volatilidad, Ait Sahalia (1996) estima la función de densidad marginal de la tasa de interés

$$\pi(r) = \frac{\sum_{i=1}^n K((r - r_i)/h)}{nh} \quad (3.36)$$

mediante el método del kernel con K la función kernel y h el parámetro de suavizado o ancho de banda correspondiente. Respecto a la tendencia, supone que existe reversión lineal a la media $\alpha(r) = \beta(m - r)$.

Posteriormente, Stanton (1997) realiza una estimación totalmente no paramétrica de la tendencia y la volatilidad del proceso estocástico de la tasa de interés a partir de los datos observados, ya que de la misma forma que no existe evidencia de cuál es el comportamiento de la volatilidad del proceso, tampoco existe para la tendencia. El procedimiento consiste en lo siguiente; como la función de densidad de transición no existe para la estimación no paramétrica, Stanton (1997) construye, en primer lugar, una familia de aproximaciones para la tendencia y la difusión del proceso con un orden de convergencia k , con k un entero positivo. Las aproximaciones de primer orden, que se obtienen mediante la discretización de Euler del proceso estocástico, son las siguientes

$$\alpha(r_t) = \frac{1}{\Delta} \mathbb{E}_t[r_{t+\Delta} - r_t | r_t = r] + O(\Delta) \quad (3.37)$$

$$\rho(r_t)^2 = \frac{1}{\Delta} \mathbb{E}_t[(r_{t+\Delta} - r_t)^2 | r_t = r] + O(\Delta) \quad (3.38)$$

Éstas son las que se utilizan habitualmente en la literatura. Stanton (1997) utiliza las diferentes aproximaciones en modelos clásicos paramétricos, y encuentra que, con datos

diarios, incluso las de primer orden son casi indistinguibles de las funciones exactas. Además Fan (2003), y Fan y Zhang (2003) demuestran teóricamente y mediante simulaciones empíricas que los momentos de mayor orden no deberían ser utilizados salvo que el intervalo muestral sea muy elevado, por ejemplo dos años, ya que el valor de los factores de inflación de la varianza aumentan considerablemente al aumentar el orden de la aproximación.

Una vez obtenidas las aproximaciones (3.37) y (3.38), Stanton (1997) utiliza el método del kernel para su estimación. Considerando el estimador de Nadaraya - Watson (3.28), se obtiene

$$\mathbb{E}[r_{t+\Delta} - r_t | r_t = r] = \frac{\sum_{t=1}^{N-1} (r_{t+\Delta} - r_t) K[(r - r_t)/h]}{\sum_{t=1}^{N-1} K(r_{t+\Delta} - r_t)/h} \quad (3.39)$$

$$\mathbb{E}[(r_{t+\Delta} - r_t)^2 | r_t = r] = \frac{\sum_{t=1}^{N-1} (r_{t+\Delta} - r_t)^2 K[(r - r_t)/h]}{\sum_{t=1}^{N-1} K(r_{t+\Delta} - r_t)/h} \quad (3.40)$$

Dinámica de la curva de rendimiento en México: un análisis de componentes principales.

Para analizar la curva de rendimiento, el enfoque tradicional, introducido por Robert Litterman y José Scheinkman (1991), consiste en aplicar el análisis de componentes principales con la idea de identificar *factores comunes* o *componentes principales* que afectan los rendimientos a diferentes plazos de manera conjunta y que puedan asociarse a las propiedades de la curva completa. De acuerdo a estos autores las propiedades de cualquier curva de rendimiento pueden resumirse en tres medidas: nivel, pendiente y curvatura.

4.1. Análisis de componentes principales

4.1.0.1. Análisis factorial vs análisis de componentes principales

El análisis factorial y el análisis de componentes principales están muy relacionados. Algunos autores consideran el segundo como una etapa del primero y otros los consideran como técnicas diferentes. El análisis de componentes principales trata de hallar componentes (factores) que sucesivamente expliquen la mayor parte de la varianza total. Por su parte el análisis factorial busca factores que expliquen la mayor parte de la varianza común. En el análisis factorial se distingue entre varianza común y varianza única, en el análisis de componentes principales no se hace esa distinción se centra en la varianza total.

Mientras que el análisis de componentes principales busca hallar combinaciones lineales de las variables originales que expliquen la mayor parte de la variación total, el análisis factorial pretende hallar un nuevo conjunto de variables, menor en número que las variables originales y que exprese lo que es común a esas variables.

El análisis factorial supone que existe un factor común subyacente a todas las variables, el análisis de componentes principales no hace tal supuesto. En el análisis de componentes principales, el primer factor o componente es aquel que explica una mayor cantidad de la varianza total, el segundo factor sería aquel que explica la mayor parte de la varianza restante, es decir, de la que no explicaba el primero y así sucesivamente. De este modo es posible obtener tantos componentes como variables originales aunque esto en la práctica no tiene sentido. Algunas otras diferencias se resumen en la siguiente tabla:

Análisis de componentes principales	Análisis de factores
No se asume ningún modelo	Existe un modelo definido
Se concentra en la diagonal de la matriz de covarianza	Se concentra en los elementos fuera de la diagonal de la matriz de covarianzas
Se puede tener un componente principal que corresponde a una sola variable	Cada factor debe contribuir por lo menos a dos variables
Un aumento en el número de componentes principales utilizados no altera a los componentes principales originales.	Un aumento en el número de factores modificación a los factores originales
Las componentes principales son combinaciones lineales de los datos originales	Los factores no son funciones lineales de los datos originales.

Tabla 4.1: Diferencias entre el análisis de componentes principales y el análisis de factores.

4.1.0.2. Desarrollo del análisis de componentes principales

El análisis de componentes principales (ACP de aquí en adelante) es un método estadístico multivariado que describe el comportamiento de variables aleatorias correlacionadas en términos de un pequeño número de variables no correlacionadas llamadas componentes principales. La idea subyacente es que la dinámica de las variables originales puede ser

descrita por un número pequeño de estas componentes.

Consideremos un modelo con N variables aleatorias. Estas variables aleatorias se incluyen en un vector columna \mathbf{X} . Para efectos de notación se asume que estas variables aleatorias están centradas. El objetivo de ACP es describir los datos que tenemos con $K < N$ variables aleatorias ortogonales, llamadas *componentes principales*, las cuáles serán combinaciones lineales de las variables aleatorias originales. Denotamos la k -ésima componente principal como:

$$Y_k = \mathbf{X}^T \mathbf{w}_k \text{ para } k = 1, \dots, N \quad (4.1)$$

donde los vectores de peso, satisfacen el siguiente problema de maximización:

$$\max_{\mathbf{w}_k \in \mathbb{R}^N} \text{Var}(\mathbf{X}^T \mathbf{w}_k) \quad (4.2)$$

Hemos de maximizar la varianza de cada componente principal, de modo que cada componente describa la mayor parte de la variabilidad total. La restricción de que cada vector de peso debe tener una longitud igual a 1 sólo sirve para quitar la indeterminación, por lo tanto los vectores \mathbf{w}_k forman un sistema ortonormal. De hecho, es fácil demostrar que (4.2) se resuelve tomando \mathbf{w}_k como el k -ésimo vector propio, es decir, el vector propio asociado con k -ésimo valor propio λ_k más grande, de la matriz de varianzas-covarianzas Σ . La varianza de la k -ésima componente principal es por lo tanto igual a:

$$\text{Var}(\mathbf{X}^T \mathbf{w}_k) = \mathbf{w}_k^T \Sigma \mathbf{w}_k = \lambda_k \quad (4.3)$$

En el ACP la proporción de varianza explicada por el k -ésimo factor o componente principal se calcula mediante la siguiente relación:

$$\frac{\text{Var}(Y_k)}{\text{Var}(Y_1 + Y_2 + \dots + Y_N)} = \frac{\lambda_k}{\lambda_1 + \lambda_2 + \dots + \lambda_N} \quad (4.4)$$

Note que todos los valores propios de una matriz de varianzas-covarianzas (o correlación) obtenida a partir de los datos son positivos, ya que cualquier matriz de varian-

covarianza propia es definida positiva¹.

Ahora denotemos la descomposición espectral de la matriz de varianzas-covarianzas como:

$$\Sigma = \mathbf{W}\Lambda\mathbf{W}^T$$

Entonces, es evidente que podemos escribir $\mathbf{Y} = \mathbf{W}^T\mathbf{X}$, y de esta relación se puede obtener $\mathbf{X} = \mathbf{W}\mathbf{Y}$.

En general, se utilizan $K < N$ componentes principales, de modo que tendremos:

$$\mathbf{Y}_{(k)} = \mathbf{W}_{(k)}^T\mathbf{X} \quad (4.5)$$

donde el subíndice (K) indica que estamos utilizando sólo los primeros k componentes principales. No podemos hallar directamente la inversa de esta relación, pero si hacemos la regresión de $\mathbf{Y}_{(k)}$ sobre \mathbf{X} , podemos demostrar que el estimador de mínimos cuadrados ordinarios de \mathbf{A} en $\mathbf{X} = \mathbf{A}\mathbf{Y}_{(k)} + \epsilon$ es $\mathbf{W}_{(k)}$, de modo que finalmente tenemos:

$$\mathbf{X} = \mathbf{W}_{(k)}\mathbf{Y}_{(k)} + \epsilon \quad (4.6)$$

Como nota final, sabemos que por definición de (4.2) la j -ésima entrada de un vector de peso \mathbf{w}_k contiene el peso con el cual \mathbf{X}_j está incrustado dentro del k -ésimo componente principal. Dentro del ACP, los vectores propios re-escalados $\sqrt{\lambda_k}\mathbf{w}_k$ son llamados *factores*, y sus entradas se les denomina *cargas factoriales*.

La decisión de mantener sólo las primeras r componentes principales nos permite reemplazar la relación exacta anterior por una aproximación:

$$x_i \approx \beta_i\mathbf{y} = \beta_{i1}y_1 + \beta_{i2}y_2 + \cdots + \beta_{ir}y_r$$

¹Hay situaciones en las que se puede obtener una estimación de una matriz de varianzas-covarianzas que no es definida positiva, por ejemplo, cuando se tienen datos perdidos de una o más de las variables observadas. Sin embargo, cualquier matriz de varianzas-covarianzas propia debe ser definida positiva, pues de lo contrario, podemos construir una combinación lineal de nuestras variables aleatorias que tenga una variación negativa. Evidentemente, esto no puede ser.

4.1.1. Revisión de la literatura

La literatura dedicada al estudio de factores comunes en la estructura temporal es bastante extensa, pudiéndose dividir en dos grandes bloques. En un primer bloque se engloban aquellos trabajos que utilizan diversas técnicas estadísticas y econométricas para cuantificar directamente cuántas variables son necesarias para caracterizar la dinámica de la ETTI. En función de la metodología utilizada podemos dividir este primer bloque en dos subgrupos.

En el primero de ellos se consideran aquellos trabajos que utilizan técnicas de componentes principales para determinar cuántas variables son necesarias para explicar un porcentaje suficientemente elevado de las fluctuaciones en los rendimientos de los bonos a distintos vencimientos y/o las variaciones de las tasas cupón cero.

Dentro de este grupo se encuentra el trabajo de Litterman y Scheinkman(LS), quienes fueron los primeros en utilizar el análisis de componentes principales para realizar este tipo de tareas. Ellos encontraron que más del 98 % de la variación en los rendimientos de los instrumentos de renta fija en Estados Unidos se puede explicar en función de tres factores o atributos de la curva de rendimiento llamados: **nivel**, **pendiente** y **curvatura**.

Posteriormente, en 1996, Knez, Litterman y Scheinkman (1996) utilizaron la misma técnica, con el fin de desarrollar modelos de factores que influyen en el rendimiento de los bonos. Analizaron los rendimientos semanales de 38 instrumentos del mercado monetario con vencimientos de 1 a 12 meses, que representan cinco diferentes segmentos de riesgo en los Estados Unidos para el período desde enero 1985 hasta agosto 1988. Y fueron capaces de identificar e interpretar al menos tres factores. El primero, que llamaron **nivel**, corresponde a los desplazamientos de la curva de rendimiento, este componente explicó el 62 % de la variación total en los rendimientos. El segundo, que denominaron **stepness**, que equivale a la pendiente de la curva, explica el 11 % de los cambios en los rendimientos. Por último, el **tesoro o curvatura** llamado así porque, como señalan su efecto es aumentar la curvatura de la curva de rendimientos en el rango de vencimientos menores a 20 años, el poder explicativo del tercer factor es aproximadamente del 13 %. Como puede deducirse directamente, los tres factores representan el 86 % de la variación en los rendimientos del

mercado monetario. Estas etiquetas representan el efecto que los factores tienen en la curva de rendimiento en lugar de describir las fuentes económicas de los choques.

En esta clase también se ubican las investigaciones de Barber y Copper(BC), Bühler y Zimmermann (BZ), D' Ecclesia y Zenios (DZ), Golub y Tilman (GT), Kanony y Mokrane (KM), Kärki y Reyes (KR), Lardic, Priaulet y Priaulet (LPP), Lekkos (L), Martellini y Priaulet (MP). Resumimos sus principales resultados en la siguiente tabla.

Autores	País(periodo) - tipo de tasa	Rango	Factores	% explicado
LS(1991)	USA (1984-1988)-spot	6M-18A	3	88.04/8.38/1.97
KM(1992)	Francia (1989-1990)-spot	1A-25A	2	93.7/6.1
DZ(1994)	Italia (1988-1992)-spot	6M-7A	3	93.91/5.49/0.42
KR(1994)	Alemania/USA (1990-1994)-spot	3M-10A	3	Total: 97/98/98
BC(1996)	USA (1985-1991)-spot	1M-20A	3	80.93/11.85/4.36
BZ(1996)	Alemania/Suiza (1988-1996)-spot	1M-10A	3	71/18/4; 75/16/3
GT(1997)	USA (1984-1995)-spot	3M-30A	3	92.8/4.8/1.27
L(2000)	Inglaterra(1987-1995)-forward	1A-9A	5	63.5/6.3/7.5/8.1/5.3
MP(2000)	Francia (1995-1998)-spot	1M-10Y	3	66.64/20.52/6.96
LPP(2003)	Alemania/Italia (1998-2000)-spot	1M-30A	3	61/23/6; 59/24/7

Tabla 4.2: Estudios sobre el ACP y el análisis factorial en mercados extranjeros.

En particular, el estudio de Lardic (2003) *et al.* cuestiona las diferentes metodologías que se pueden adoptar cuando se usa el ACP. Podemos preguntarnos si los resultados dependen de la elección de la metodología.

- ¿Debemos utilizar los niveles de las tasas de interés o los cambios en las tasas de interés?. Por ejemplo, Kanony y Mokrane (1992) utilizan los niveles de las tasas de interés, mientras que otros autores prefieren considerar los cambios.

- ¿Usar la matriz de correlación o la matriz de varianzas-covarianzas?. Por ejemplo, Barber y Copper(1996) utilizan la matriz de varianzas-covarianzas y Buhler y Zimmerman(1996) utilizan la matriz de correlación.
- Los resultados del ACP dependen del número de entradas y vencimientos, por ejemplo, Golub y Tilman(1997) consideran todo el espectro de vencimientos (utilizan 10 variables, con vencimientos que van desde los 3 meses hasta 30 años), mientras que D'Ecclesia y Zenios(1994), tomar solamente los segmentos de corto y mediano plazo (8 variables, con vencimientos que van de 6 meses a 7 años).
- Si los resultados del ACP dependen de la frecuencia de los datos, se deben utilizar datos diarios, semanales o mensuales y durante qué período de tiempo. Por ejemplo, Barber y Copper(1996) toman datos mensuales desde agosto de 1985 hasta febrero de 1991 así como D'Ecclesia y Zenios (1994) consideran datos semanales de 1988 a 1992.

Lardic *et al.*(2003) concluyen que el ACP debe ser implementado considerando los cambios en las tasas de interés ya que son estacionarios, y segundo, que estas variables deben estar centradas por cuestiones de reducción de la varianza. El número de variables y los plazos que cubren puede modificar de manera significativa el conjunto de datos y el porcentaje de varianza explicada por los factores, y que la exactitud de los resultados es proporcional a la frecuencia de los datos.

Para el caso particular de México se tienen los siguientes estudios:

- Jiménez (2002), en su trabajo *Testing the Stability of the Components Explaining Changes of the Yield Curve in Mexico. A Principal Component Analysis Approach*, estima los componentes principales que determinan los movimientos de la curva de rendimientos en el mercado mexicano, al tiempo que busca observar su estabilidad a través del tiempo. La hipótesis central era que mientras los componentes principales son muy estables o permanecen sin cambios, la varianza explicada por ellos fluctúa sustancialmente. El trabajo concluye que los tres componentes de la curva

de rendimientos nivel, pendiente y curvatura son relevantes a la hora de explicar los desplazamientos totales de la estructura a plazos de la economía mexicana.

- Cortés, Ramos y Torres (2009), estudian la dinámica de la estructura de plazos de las tasas de interés en México del 2001 al 2008, y hallaron que más del 99% de la variación total en la curva de rendimiento se puede explicar por tres factores: el nivel, la pendiente y la curvatura. También muestran que el factor nivel se correlaciona positivamente con medidas de inflación a largo plazo y que el factor pendiente se relaciona negativamente con la tasa de interés a un día. Finalmente concluyen que la estructura temporal de las tasas de interés en México, a pesar de su relativa corta existencia, se comporta como en los mercados que tienen sistemas financieros más desarrollados.
- Siguiendo con la literatura existente aplicada a México, Villarreal(2001) encontró que las tasas de interés en México son una serie no estacionaria y que el proceso Random Walk no es la mejor representación para estimar y predecir las tasas de interés a corto plazo. Por ende, la existencia de la no-estacionariedad de los datos establece la necesidad de usar las primeras diferencias de las tasas de interés en lugar de los niveles para obtener inferencias correctas. El autor además encontró que pruebas a la hipótesis de las expectativas revelan que la parte más corta de la estructura de plazos mexicana (i.e. menos de seis meses) da información sobre la tasa de interés real e inflación futura.

Un segundo grupo de trabajos lo constituyen aquellos que utilizan técnicas de cointegración para determinar el número de factores comunes no estacionarios presentes en la estructura temporal de las tasas de interés en distintos mercados. En general los resultados obtenidos en estos trabajos apuntan a que hay un único factor común no estacionario en las tasas de interés inferiores a 1 año (y esto es independiente del mercado que se analice) y tendencias adicionales en las tasas a medio y largo plazo.

Un segundo bloque de literatura, orientado a la administración de portafolios de títulos de deuda pública, se ha desarrollado recientemente a partir del trabajo pionero de Elton,

Gruber y Michaely (EGM)(1990). Estos autores proponen un modelo de factores de la ETTI basado en el análisis de regresión. En este contexto diseñan una metodología para determinar que las tasas de interés, explican de forma más adecuada el comportamiento de la estructura temporal. Sobre la base de dicha metodología comparan la capacidad predictiva de un modelo bifactorial sobre un modelo unifactorial para datos del mercado de deuda de EEUU. Los resultados obtenidos indican que la capacidad para replicar el comportamiento de la ETTI es mayor en el modelo con dos factores que en el que considera una única variable.

4.2. Metodología

En esta sección se incluye la descripción y el análisis estadístico de los datos. De la misma manera, se presentan los resultados del análisis de componentes principales.

4.2.1. Datos

Los instrumentos que generalmente integran la curva de rendimiento en México son los CETES (bonos cupón cero emitidos por primera vez en 1978) de 28, 91, 182 y 364 días; así como bonos gubernamentales con cupón semestral de 3 años, 5 años (su primera emisión se llevó a cabo en el 2000), 7 años (emitidos en 2002), 10 años (emitidos en 2001), 20 años (emitidos en 2003) y 30 años (emitidos en 2006). Para obtener las componentes principales de la ETTI en México, usamos *observaciones diarias* de los CETES publicados por Reuters para los siguientes vencimientos: 1, 28, 91, 182, 364, 728, 1092, 1820, 2548 y 3600 días, que corresponden a los plazos de 1 día, 1, 3 y 6 meses, y 1, 2, 3, 5, 7 y 10 años. El período muestral analizado se extiende desde el 9 de julio de 2002 hasta el 7 de diciembre de 2009, lo que recoge una muestra de 1867 observaciones para cada plazo al vencimiento de los CETES. Usamos esta muestra por dos razones principales. La primera es que parece razonable suponer que la inflación en México siguió un proceso estacionario durante este período. La segunda razón es que el gobierno mexicano fue capaz de emitir bonos a tasa fija para horizontes de largo plazo (10 años) desde 2001. Las figuras (4.1) y (4.2) ilustran la evolución del rendimiento de los CETES a 1, 28, 91, 182, 364, 728, 1092,

1820, 2548 y 3600 días en el período muestra considerado.

Como se puede observar, en general el nivel del rendimiento de los bonos disminuye sobre el horizonte temporal. Además, el rendimiento es menor a corto plazo que a largo plazo, en este sentido, es posible considerar un premio por madurez² que se encuentra relacionado con el plazo de vencimiento, entre mayor sea el periodo de madurez de un instrumento se espera que mayor sea la tasa que paga éste con el fin de compensar las posibles fluctuaciones del entorno financiero. También, la pendiente de la curva de rendimiento se reduce.

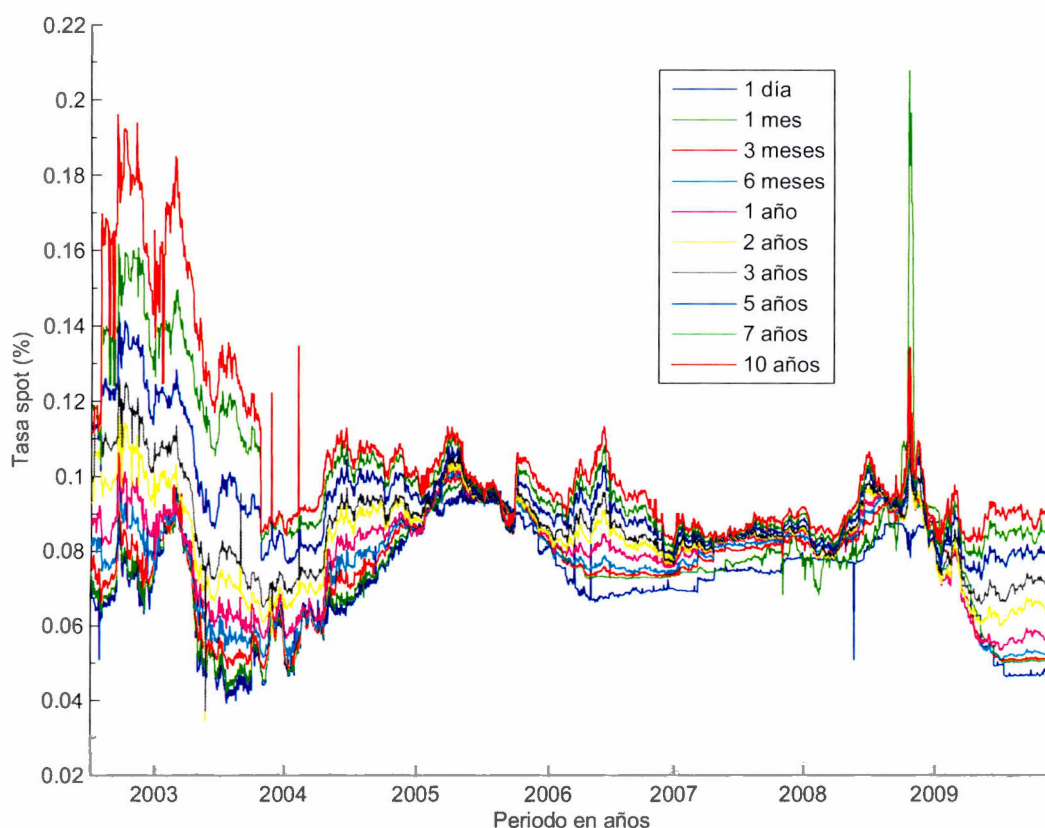


Figura 4.1: Observaciones diarias CETES: julio 2002 a diciembre 2009

²Estos resultados apoyan la teoría de la preferencia por la liquidez de Hicks(1946), donde considera la existencia de un premio por madurez positivo (este premio es considerado también como un premio por riesgo o una prima por liquidez).

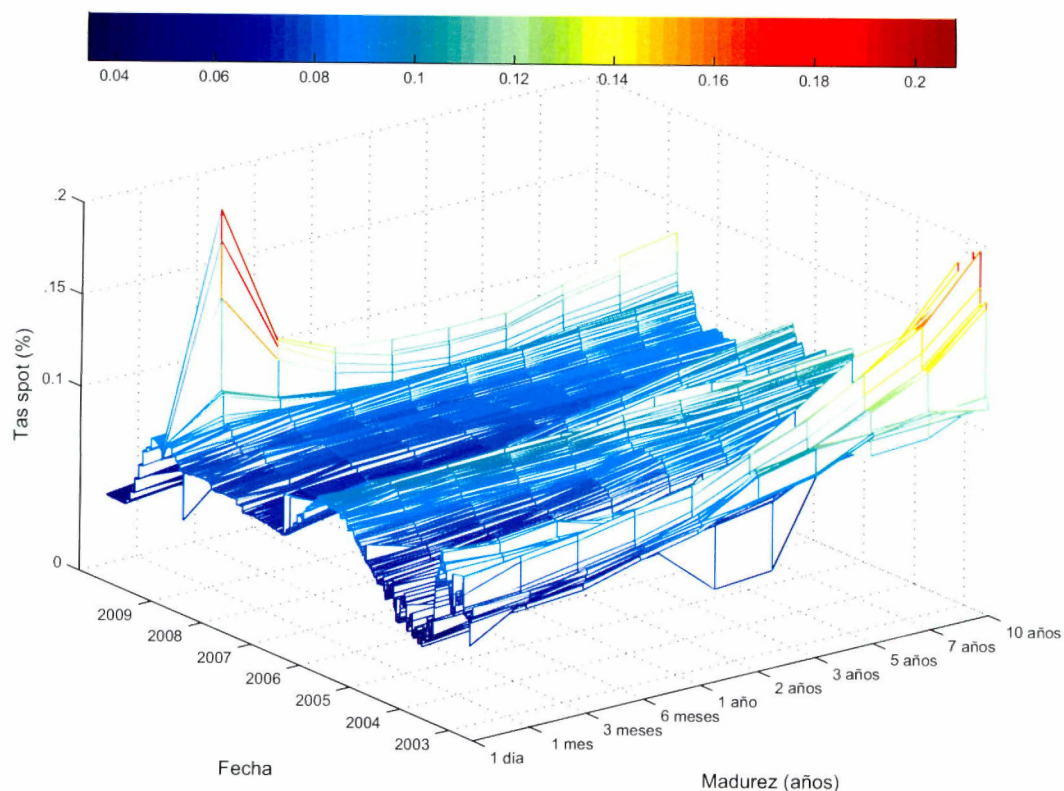


Figura 4.2: Rendimientos de los CETES: julio 2002 a diciembre 2009

4.2.2. Estadística descriptiva: niveles

Antes de llevar a cabo el análisis de componentes principales de los bonos cupón cero, vale la pena ilustrar el comportamiento de la muestra completa. La siguiente tabla muestra las estadísticas descriptivas de los rendimientos de las tasas de interés para la muestra seleccionada.

Los datos se caracterizan por algunos hechos estilizados estándar. La media de los 10 instrumentos aumenta gradualmente en relación a la madurez de éstos confirmando la presencia de una prima. Con respecto a la desviación estándar de los rendimientos éstas disminuyen con la madurez en un primer momento, pero luego aumentan con la madurez como se muestra en la figura 4.3. Este comportamiento es contrario a lo que se ha visto en hechos estilizados:³ las tasas de largo plazo son menos volátiles y más persistentes que las

³Diebol y Li (2003) apuntan esta situación. Generalmente las tasas de interés de corto plazo son más volátiles que las de largo plazo debido a que estas últimas reflejan expectativas y las de corto plazo

Madurez	N	Media	Desv. típ.	Asimetría	Curtosis
Un día	1867	.07214	.01376	-.456	-.529
Un mes	1867	.07514	.01613	1.279	9.546
Tres meses	1867	.07632	.01410	-.299	-.219
Seis meses	1867	.07766	.01313	-.497	-.469
Un año	1867	.07984	.01221	-.551	-.530
Dos años	1867	.08364	.01166	-.152	-.375
Tres años	1867	.08685	.01177	.508	.331
Cinco años	1867	.09301	.01314	1.410	1.989
Siete años	1867	.09865	.01732	1.650	2.345
Diez años	1867	.10399	.02354	1.974	3.291

Tabla 4.3: Estadística descriptiva: niveles

de corto plazo. Los niveles muestran un curtosis similares a tres a largo plazo, asimetría negativa a mediano plazo, es decir, una mayor concentración de valores a la izquierda de la media y sesgo positivo a largo plazo.

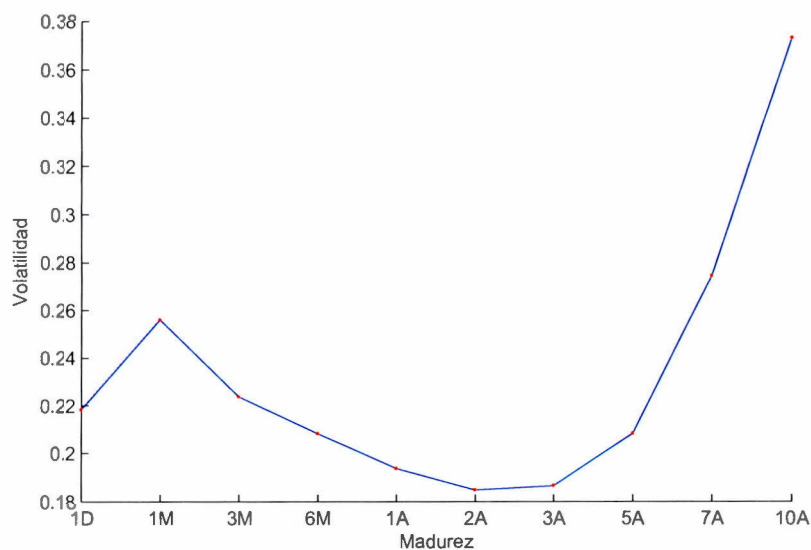


Figura 4.3: Volatilidad anualizada de los CETES en función de los distintos vencimientos.

aumentan factores coyunturales.

4.2.3. Estadísticas descriptivas: diferencias

Las primeras diferencias (cambios o rendimientos diarios) en el nivel y forma de la curva de rendimiento conduce el riesgo a corto plazo y al comportamiento de los bonos cupón cero. Como un bono cupón cero no tiene pagos de intereses, su retorno es impulsado por los cambios de los precios. Estos cambios pueden surgir por dos fuentes. La primera es la simple acumulación de los precios hacia el valor de la madurez que ocurre en el tiempo (los bonos cupón cero se emiten al descuento y maduran a la par). La segunda fuente es un cambio en el rendimiento. En un horizonte de tiempo relativamente corto, la segunda fuente es por mucho la más importante.

Madurez	N	Media	Desv. típ.	Asimetría	Curtosis
Un día	1866	-1.3704E-05	0.00206894	-0.114	52.071
Un mes	1866	-1.2296E-05	0.00228867	5.812	189.64
Tres meses	1866	-1.2799E-05	0.0013661	1.755	44.013
Seis meses	1866	-1.5069E-05	0.00118633	0.922	16.669
Un año	1866	-1.5353E-05	0.00128383	-0.498	15.81
Dos años	1866	-1.4418E-05	0.00198543	-0.193	139.622
Tres años	1866	-1.4377E-05	0.00198519	-0.185	161.358
Cinco años	1866	-1.7392E-05	0.00173141	0.259	81.669
Siete años	1866	-1.1755E-05	0.00257883	0.221	109.5
Diez años	1866	-1.1893E-05	0.00370983	1.079	74.455

Tabla 4.4: Estadística descriptiva: diferencias.

En la tabla 4.4 se presentan algunas estadísticas descriptivas de las primeras diferencias o cambios diarios de los rendimientos durante el período muestra considerado. Tres observaciones importantes pueden ser hechas de estos resultados. En primer lugar, como es de esperarse, el cambio promedio en los rendimientos de las diferentes es muy pequeño, prácticamente nulo para todos los vencimientos. Teniendo en cuenta que estos representan los cambios diarios, el tamaño pequeño era de esperarse. En segundo lugar, la incertidumbre que rodea a cada uno de los promedios es muy alta, con desviaciones estándar muy grandes en relación al valor medio. En tercer lugar, la distribución de los cambios en el rendimiento claramente no es normal, por el contrario, las distribuciones son muy leptocúrticas. Finalmente, la siguiente tabla muestra la matriz de correlaciones de las primeras

diferencias.

	1D	1M	3M	6M	1A	2A	3A	5A	7A	10A
1D	1									
1M	.186	1								
3M	.231	.705	1							
6M	.214	.499	.657	1						
1A	.188	.337	.505	.714	1					
2A	.048	.154	.220	.348	.574	1				
3A	.048	.138	.193	.292	.463	.936	1			
5A	.107	.163	.217	.255	.313	.203	.349	1		
7A	.073	.125	.141	.180	.234	.157	.270	.872	1	
10A	.048	.076	.089	.182	.222	.154	.232	.639	.825	1

Tabla 4.5: Matriz de correlaciones: cambios diarios

4.2.4. Condiciones para aplicar el análisis de componentes principales.

Cabe destacar algunas cuestiones metodológicas que orientarán la implementación del análisis de componentes principales a la muestra de datos.

1. Antes de aplicar el análisis de componentes principales debe comprobarse si es necesario, es decir, si las correlaciones entre las variables analizadas son lo suficientemente grandes como para justificar la factorización de la matriz de correlación. De acuerdo a Barber y Copper (1996) para que el ACP sea fructífero es conveniente que la matriz de correlación contenga grupos de variables que correlacionan fuertemente entre sí. Una matriz de correlación próxima a una matriz identidad indica que el ACP conducirá a una solución deficiente. Para formarnos una idea sobre el grado de relación existente entre las series de observaciones diarias de los bonos cupón cero para los 10 vencimientos considerados, la tabla 4.6 ofrece, además de las correlaciones, el nivel crítico unilateral (*Sig. unilateral*) asociado a cada coeficiente de correlación.

Un nivel crítico menor que 0.05 indica que la correlación poblacional entre el correspondiente par de series de datos puede ser considerado significativamente distinta

	1D	1M	3M	6M	1A	2A	3A	5A	7A	10A	
Correlación	1D	1									
	1M	.878	1								
	3M	.956	.948	1							
	6M	.928	.895	.981	1						
	1A	.866	.827	.934	.980	1					
	2A	.729	.692	.804	.881	.946	1				
	3A	.589	.561	.664	.755	.845	.967	1			
	5A	.294	.290	.366	.471	.587	.779	.902	1		
	7A	.077	.083	.138	.243	.363	.583	.750	.955	1	
	10A	.010	.013	.058	.156	.273	.500	.683	.913	.985	1
Sig. (Unilateral)	1D										
	1M	.000									
	3M	.000	.000								
	6M	.000	.000	.000							
	1A	.000	.000	.000	.000						
	2A	.000	.000	.000	.000	.000					
	3A	.000	.000	.000	.000	.000	.000				
	5A	.000	.000	.000	.000	.000	.000	.000			
	7A	.000	.000	.000	.000	.000	.000	.000	.000		
	10A	.334	.285	.006	.000	.000	.000	.000	.000	.000	

Tabla 4.6: Matriz de correlación: niveles

de cero. Lo deseable, por tanto, es encontrar muchos niveles críticos pequeños (muchos ceros). Como se puede observar, la correlación es más alta cuanto menor es la diferencia temporal entre los vencimientos (las correlaciones tienden a decrecer con el spread). Alta correlación a corto plazo, media a media, y baja a largo plazo. Por último el valor del *determinante* de la matriz de correlación vale $2.31E - 014$, lo que indica que las series de observaciones diarias de los CETES para los distintos vencimientos están linealmente relacionadas (son linealmente dependientes), lo cual es un buen síntoma de cara a la idoneidad del ACP.

Cuando las variables originales están altamente correlacionadas, la mayor parte de su variabilidad puede ser explicada por menos componentes. La afirmación de que las correlaciones son lo suficientemente fuertes para justificar la aplicación del ACP puede ser avalado mediante la realización de dos pruebas estadísticas, a saber, la

prueba de esfericidad de Bartlett y la medida de Kaiser-Meyer-Olkin (KMO).

La prueba de esfericidad de Bartlett (1950), parte de la hipótesis nula de que la matriz de coeficientes de correlación \mathbf{R} no es significativamente distinta de la matriz identidad⁴, esto es:

$$H_0 : \mathbf{R} = \mathbf{I}$$

Bartlett calcula el estadístico de prueba basado en el valor del determinante de la matriz de correlación del siguiente modo $V = |\mathbf{R}|$ y se rechaza H_0 si:

$$- \left[n - 1 - \frac{(2p + 5)}{6} \right] \ln |V| > \chi_{\alpha, p(p-1)/2}^2$$

donde n es el número de observaciones diarias por vencimiento y p el número de variables o series de datos. Si se rechaza H_0 se recomienda realizar el análisis de componentes principales.

La medida de adecuación muestral de Kaiser-Meyer-Olkin (KMO) es un índice que compara la magnitud de los coeficientes de correlación observados con la magnitud de los coeficientes de correlación parcial y de esta manera crea un índice que va de 0 a 1, llegando a 1 cuando cada variable se puede predecir con exactitud por los demás (sin errores).

Los resultados de la prueba de Bartlett y medida KMO son:

Medida de adecuación muestral de Kaiser-Meyer-Olkin.		.757
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	58458.770
	gl	45
	Sig.	.000

Tabla 4.7: KMO y prueba de Bartlett.

Si el valor de la medida de adecuación muestral (KMO) es pequeño (los valores por debajo de 0.6 se consideran mediocres) puede que no sea pertinente utilizar el

⁴Que equivale a decir que la matriz de varianza-covarianza Σ es diagonal

ACP. Si el nivel crítico (*Sig.*) en la prueba de esfericidad de Bartlett es mayor que 0.05 no podemos rechazar la hipótesis nula de esfericidad y, consecuentemente, no podemos asegurar que el ACP sea adecuado para la muestra. Los resultados de la tabla 4.7 siguieren que es posible llevar a cabo el análisis de componentes principales.

2. Lardic, Priault y Priault (2001) reportan que los niveles (tasas brutas) de las tasas de interés a los diferentes plazos tienden a estar más correlacionados que los cambios en las tasas (primeras diferencias) hecho que se puede verificar en nuestro caso al comparar los registros de las tablas 4.5 y 4.6. Esta diferencia de correlaciones está ligada a la presencia de una tendencia determinista o estocástica en las variables en cuestión, lo que hace que las series de los niveles (tasas brutas) sean no estacionarias, por lo que es necesario utilizar la primera diferencia de las observaciones. Para comprobar la validez de este argumento para el caso de la ETTI en México vamos a aplicar las pruebas de estacionariedad de Dickey-Fuller Aumentada (ADF); Phillips-Perron (PP); y la prueba de Kwiatkowski, Phillips, Schmidt y Shin (KPSS) a los niveles y primeras diferencias de las tasas para los diferentes vencimientos.

La prueba de Dickey-Fuller aumentada tiene como hipótesis nula la existencia de una raíz unitaria, es decir, que la variable analizada es no estacionaria. La hipótesis nula se rechaza si el estadístico de prueba ADF es menor que el valor crítico t^* . Sin embargo, aunque es ampliamente utilizada, su poder explicativo es limitado, como lo demuestra Pindyck y Rubinfeld (2004), ya que sólo permite rechazar la hipótesis de que la variable no es una caminata aleatoria.

La prueba de Phillips-Perron, a su vez, también considera la existencia de una raíz unitaria, la cual se rechaza si el estadístico de prueba PP es menor que el valor crítico t^* . Sin embargo al ser una prueba no paramétrica presenta estadísticos más robustos que los de (DFA) (Lardic, Priault, Priault (2001)).

Finalmente, la prueba de Kwaiatkowski, Phillips, Schmidt y Shin (KPSS) tiene como hipótesis nula la estacionariedad de la serie (la no existencia de una raíz unitaria) y se rechaza, si el estadístico KPSS es mayor que el valor crítico t^* . Esta prueba ha sido ampliamente utilizada en la literatura en lugar de pruebas cuya hipótesis nula es la existencia de una raíz unitaria.

Las tablas 4.8 y 4.9 presentan los resultados obtenidos al aplicar las pruebas anteriores.

Vencimiento	Variables en niveles			Variables en primera diferencia		
	Estadístico de prueba			Estadístico de prueba		
	ADF	PP	KPSS	ADF	PP	KPSS
Un día	-0.8431 **	-0.7561 **	0.4580	-60.089 *	-64.791 *	0.129 *
Un mes	-0.8331 **	-0.9593 **	0.5096	-16.920 *	-35.86 *	0.0310 *
Tres meses	-0.7113 **	-0.7133 **	0.4632	-45.273 *	-45.235 *	0.1130 *
Seis meses	-0.7949 **	-0.7949 **	0.4125	-43.776 *	-43.773 *	0.1267 *
Un año	-0.8085 **	-0.8365 **	0.4610	-48.654 *	-48.819 *	0.1051 *
Dos años	-0.8038 **	-0.8045 **	0.7441 **	-61.73 *	-66.804 *	0.0612 *
Tres años	-0.7947 **	-0.7947 **	1.1271 **	-59.867 *	-63.526 *	0.0415 *
Cinco años	-0.8448 **	-0.8835 **	2.3210 **	-51.306 *	-52.192 *	0.0346 *
Siete años	-0.8527 **	-0.8226 **	2.9558 **	-57.663 *	-61.864 *	0.0344 *
Diez años	-0.8823 **	-0.8226 **	2.8267 **	-23.811 *	-61.87 *	0.0442 *
* Se rechaza Ho: No estacionaria			** No se rechaza Ho: No estacionaria			

Tabla 4.8: Pruebas de estacionariedad para los niveles y primeras diferencias de los CETES.

Nivel de significancia	ADF	PP	KPSS
	Valores críticos t^*		
1 %	-2.566202	-2.566202	0.739
5 %	-1.940993	-1.940993	0.463
10 %	-1.616585	-1.616585	0.347

Tabla 4.9: Valores críticos para las pruebas de estacionariedad.

Para las tres pruebas ADF,PP y KPSS, la hipótesis de estacionariedad se rechaza para las tasas brutas al nivel de significancia del 1%, 5% y 10%, excepto para los vencimientos de un día a un año para la prueba KPSS a los niveles de significancia del

1 % y 5 %. En cuanto a las pruebas para la primera diferencia, los resultados muestran la existencia de estacionariedad en la serie, lo que demuestra la validez del argumento presentado por Lardic, Priaulet y Priaulet (2001), en utilizar las diferencias o los cambios diarios de las tasas de interés para el análisis de componentes principales.

3. Otro punto importante se refiere al uso de la matriz de correlación o la matriz de varianzas-covarianzas al aplicar el ACP, lo que equivale a la estandarización o no de las variables. Como lo señalan Lardic, Priaulet y Priaulet (2001), como la ETTI no es plana y existen diferencias en la volatilidad⁵ de las tasas a distintos vencimientos (ver figura 4.3), el no estandarizar las variables (uso directo de la matriz de covarianza) introduce un sesgo en el resultado. Cuando las variables están expresadas en diferentes unidades de medida se sugiere trabajar con la matriz de correlación para aislar los efectos nocivos de las distintas unidades de medida. Con la estandarización de las variables, las matrices de correlación y covarianzas **son idénticas**. Así, la crítica de usar la matrices de covarianza, enunciada por Lardic, Priaulet, Priaulet (2001) no aplican al estandarizar las variables.

4.2.5. Aplicación del ACP

4.2.5.1. Configuración general

Tenemos 10 variables y 1867 observaciones de cada una de estas variables. Consideramos los 1866 cambios en las tasas de interés diarias $\Delta R(t, \theta_k) = R(t + 1, \theta_k) - R(t, \theta_k)$, y llevamos a cabo un ACP de los datos centrados y estandarizados, lo que equivale a utilizar la matriz de correlación de los cambios $\Delta R(t, \theta_k)$ de las tasas de interés. La idea es definir la matriz

$$\Delta R = (\Delta R_{t_k})_{\substack{1 \leq t \leq 1866 \\ 1 \leq k \leq 10}} = \left(\frac{\Delta R(t, \theta_k) - \overline{\Delta R(\cdot, \theta_k)}}{\sqrt{1866} \sigma_{\Delta R(\cdot, \theta_k)}} \right)_{\substack{1 \leq t \leq 1866 \\ 1 \leq k \leq 10}}$$

⁵Pruebas de Levene para determinar si existen diferencias significativas en las varianzas de los datos para diferentes vencimientos demostrarán que en la mayoría de los casos las varianzas de los CETES a distintos vencimientos difieren al nivel de significancia del 1 % y 5 %, salvo los casos de las tasas corto plazo; y las tasas a largo plazo no difieren significativamente entre sí.

donde $\overline{\Delta R(\cdot, \theta_k)}$ y $\sigma_{\Delta R(\cdot, \theta_k)}$ son, respectivamente, el valor promedio y la desviación estándar de los cambios en la tasa de interés con madurez θ_k . Podemos escribir cada elemento de la matriz ΔR como

$$\frac{\Delta R(t, \theta_k) - \overline{\Delta R(\cdot, \theta_k)}}{\sqrt{1866} \sigma_{\Delta R(\cdot, \theta_k)}} = \sum_{l=1}^{10} \sqrt{\lambda_l} V_{ul} U_{lk} \quad (4.7)$$

donde $(U) = (U_{kl})_{1 \leq k, l \leq 10}$ es la matriz de los 10 vectores propios de la matriz de correclación $\Delta R^T \Delta R$ de los cambios de las tasas de interés,

$(U^T) = (U_{lk})_{1 \leq k, l \leq 10}$ la matriz transpuesta de U ,

$(V) = (V_{ul})_{\substack{1 \leq l \leq 1866 \\ 1 \leq u \leq 10}}$ la matriz de los 10 vectores propios de $\Delta R \Delta R^T$,

λ_l el valor propio correspondiente al vector propio U_l .

Usando la ecuación (4.7) obtenemos

$$\Delta R(t, \theta_k) = \overline{\Delta R(\cdot, \theta_k)} + \sum_{l=1}^{10} \sqrt{1866} \sigma_{\Delta R(\cdot, \theta_k)} \sqrt{\lambda_l} V_{ul} U_{lk}$$

o bien

$$\Delta R(t, \theta_k) = \overline{\Delta R(\cdot, \theta_k)} + \sum_{l=1}^{10} c_{lk} C_t^l$$

donde

$$c_{lk} = \sqrt{1866} \sigma_{\Delta R(\cdot, \theta_k)} \sqrt{\lambda_l} U_{lk}$$

$$V_{ul} = C_t^l$$

con C_t^l el l -ésimo componente principal asociado al l -ésimo factor, y c_{lk} la coordenada de la tasa de interés con vencimiento θ_k sobre el l -ésimo factor⁶. El coeficiente c_{lk} es también la sensibilidad de $\Delta R(t, \theta_k)$ con respecto al factor l , como se puede ver a través de la ecuación (4.8)

$$\frac{\Delta(\Delta R(t, \theta_k))}{\Delta C_t^l} = c_{lk} \quad (4.8)$$

⁶Note que los factores son linealmente independientes entre sí, ya que representan ejes ortogonales. Están ordenadas de manera decreciente con base a los valores propios correspondientes.

También se puede escribir

$$\Delta R(t, \theta_k) = \overline{\Delta R(\cdot, \theta_k)} + \sum_{l=1}^M c_{lk} C_t^l + \sum_{l=M+1}^{10} c_{lk} C_t^l = \overline{\Delta R(\cdot, \theta_k)} + \sum_{l=1}^M c_{lk} C_t^l + \epsilon_{lk}$$

donde M es el número de factores usados y ϵ_{lk} es el término residual el cual se asume que no contribuye a los cambios de las tasas de interés. Finalmente:

$$\Delta R(t, \theta_k) \approx \overline{\Delta R(\cdot, \theta_k)} + \sum_{l=1}^M c_{lk} C_t^l \quad (4.9)$$

4.2.5.2. Resultados.

La Tabla 4.10 muestra los valores y vectores propios asociados a cada componente principal, así como los porcentajes de varianza explicada y varianza acumulada explicada por los diez componentes principales de la matriz de correlación de los cambios diarios para el periodo comprendido del 9 de julio de 2002 hasta el 7 de diciembre de 2009.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Valor propio	6.816	2.760	0.235	0.116	0.044	0.016	0.006	0.005	0.002	0.001
% explicado	68.2	27.6	2.3	1.2	0.4	0.2	0.1	0	0	0
% acumulado	68.2	95.8	98.1	99.3	99.7	99.9	100	100	100	100
Vectores propios:										
Variable	VP1	VP2	VP3	VP4	VP5	VP6	VP7	VP8	VP9	VP10
Un día	.820	-.495	.104	.257	-.078	-.008	.017	-.002	-.002	.000
Un mes	.800	-.488	.280	-.203	-.045	.003	.020	-.004	-.005	.000
Tres meses	.883	-.458	.080	.004	.036	.001	-.047	.014	.027	.001
Seis meses	.930	-.349	-.041	.017	.095	-.003	-.020	.004	-.035	.003
Un año	.962	-.206	-.145	-.008	.098	-.004	.035	-.026	.015	-.010
Dos años	.971	.075	-.212	-.045	-.039	.030	.019	.025	.005	.020
Tres años	.930	.307	-.164	-.046	-.101	.024	-.021	-.002	-.006	-.019
Cinco años	.753	.651	.003	-.018	-.033	-.080	-.014	-.029	.001	.011
Siete años	.566	.812	.120	.022	.044	-.036	.018	.042	.001	-.009
Diez años	.489	.848	.174	.055	.034	.080	-.004	-.020	.000	.005

Tabla 4.10: Valores propios y porcentaje de varianza explicada por los cambios de las tasas de interés.

Se puede apreciar que el primer componente explica el 68.2%, el segundo el 27.6% y el

tercero el 2.3 %, en conjunto los tres primeros componentes principales explican el 98.1 % de la varianza total. Estos resultados son similares a los obtenidos en estudios desarrollados por:

- Litterman y Scheinkman (1991): en uno de los trabajos más influyentes *Common Factors Affecting Bond Returns* hallaron que los tres primeros componentes explican el 98.38 % de la varianza total. La primera componente explica el 88.03 %, la segunda componente el 8.38 % y la tercera componente principal el 1.97 %.
- Barber y Copper (1996): en su artículo *Immunisation Using Principal Component Analysis* los tres primeros factores explican el 97.15 % de la varianza total de los rendimientos de los bonos, sin embargo, la contribución a la varianza es diferente, el primer componente explica el 81 %, el segundo el 12 % y el tercero un poco más del 4 %.
- Jiménez (2002): en su trabajo *Testing the Stability of the Components Explaining Changes of the Yield Curve in Mexico. A Principal Component Analysis Approach*, estima los componentes principales que determinan los movimientos de la curva de rendimientos en el mercado mexicano en el periodo de junio de 1996 a junio de 1999, encontrando que el primer factor representa en promedio, el 95.5 % de la variación total de la curva de rendimiento, el segundo componente el 3,35 % de la variación, mientras que el tercer componente explica el 0.73 %. Por lo tanto, los tres primeros componentes principales explican el 99.6 % de la variación de las tasas de interés.
- Cortés, Ramos y Torres (2009): estudian la dinámica de la estructura de plazos de las tasas de interés en México entre el 2001 y el 2008, encontrando que el 99.31 % de la variación total de los rendimientos se puede explicar por los tres factores, el nivel que explica el 78.56 %, la pendiente que contribuye con el 16.45 % y la curvatura que explica el 4.3 % del total.

Al comparar nuestros resultados con los antes señalados es posible ver cómo, al igual que para estos últimos, el nivel y la pendiente tienen la mayor participación en lo concerniente a la explicación de la variabilidad. El 95.8 % de la varianza total de los datos

se explica a partir de los desplazamientos paralelos y de la pendiente de la curva, mientras que la curvatura explica sólo el 2.3% de los movimientos. Aun cuando los primeros dos factores concentran la explicación de los movimientos de la curva de rendimiento, la participación de cada uno de ellos da mayor peso al componente asociado con el nivel que explica alrededor del 68% de los movimientos. Lo anterior es similar a resultados de estudios tanto para mercados desarrollados como para países emergentes como México. A partir de los vectores propios de la tabla 3.10 es posible obtener un sistema de cambios diarios para los vencimientos a 1, 28, 91, 182, 364, 728, 1092, 1820, 2548 y 3600 días de la siguiente manera:

$$\begin{aligned} \Delta R(t, 1) &\approx -1.37E-05 + .82C_t^1 - .50C_t^2 + .10C_t^3 + .25C_t^4 - .08C_t^5 - .008C_t^6 + .01C_t^7 - .002C_t^8 - .002C_t^9 + 0C_t^{10} \\ \Delta R(t, 28) &\approx -1.49E-05 + .80C_t^1 - .49C_t^2 + .28C_t^3 - .20C_t^4 - .05C_t^5 + .003C_t^6 + .02C_t^7 - .004C_t^8 - .005C_t^9 + 0C_t^{10} \\ \Delta R(t, 91) &\approx -1.27E-05 + .88C_t^1 - .46C_t^2 + .08C_t^3 + .004C_t^4 + .04C_t^5 + .001C_t^6 - .04C_t^7 + .01C_t^8 + .02C_t^9 + .001C_t^{10} \\ \Delta R(t, 182) &\approx -1.50E-05 + .93C_t^1 - .35C_t^2 - .04C_t^3 + .01C_t^4 + .09C_t^5 - .003C_t^6 - .02C_t^7 + .004C_t^8 - .04C_t^9 + .003C_t^{10} \\ \Delta R(t, 364) &\approx -1.53E-05 + .96C_t^1 - .20C_t^2 - .15C_t^3 - .008C_t^4 + .10C_t^5 - .004C_t^6 + .04C_t^7 - .02C_t^8 + .01C_t^9 - .01C_t^{10} \\ \Delta R(t, 728) &\approx -1.44E-05 + .97C_t^1 + .08C_t^2 - .21C_t^3 - .05C_t^4 - .04C_t^5 + .03C_t^6 + .02C_t^7 + .02C_t^8 + .005C_t^9 + .02C_t^{10} \\ \Delta R(t, 1092) &\approx -1.43E-05 + .93C_t^1 + .30C_t^2 - .16C_t^3 - .04C_t^4 - .10C_t^5 + .02C_t^6 - .02C_t^7 - .002C_t^8 - .006C_t^9 - .02C_t^{10} \\ \Delta R(t, 1820) &\approx -1.73E-05 + .75C_t^1 + .65C_t^2 + .003C_t^3 - .01C_t^4 - .03C_t^5 - .08C_t^6 - .01C_t^7 - .03C_t^8 + .001C_t^9 + .01C_t^{10} \\ \Delta R(t, 2548) &\approx -1.75E-05 + .56C_t^1 + .81C_t^2 + .12C_t^3 + .02C_t^4 + .04C_t^5 - .03C_t^6 + .01C_t^7 + .04C_t^8 + .001C_t^9 - .009C_t^{10} \\ \Delta R(t, 3600) &\approx -1.18E-05 + .49C_t^1 + .89C_t^2 + .17C_t^3 + .06C_t^4 + .03C_t^5 - .08C_t^6 - .004C_t^7 - .02C_t^8 + 0C_t^9 + .005C_t^{10} \end{aligned}$$

4.2.5.3. Número de componentes principales a retener.

No existe un criterio único para hallar el número de componentes principales que pueden ser retenidos, ya que el poder explicativo de los componentes principales puede variar de manera arbitraria en cada caso. Sin embargo, a menudo son empleadas un par de reglas básicas.

- El criterio del *umbral* nos guía para seleccionar tantos componentes principales como sea necesario (es decir, los primeros k) mediante el grado de varianza explicada acumulada que exceda a determinado porcentaje, como por ejemplo el 90% que fue utilizado en este trabajo.

$$\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^N \lambda_i} > \text{umbral}$$

- La regla de Kaiser(1959), en ella se excluyen todos los componentes principales con valores propios menores que el promedio de todos ellos. Si se usa la matriz de correlación, o si las variables originales están estandarizadas, cada una de ellas tendrá una varianza igual a la unidad y el promedio es igual a 1. Por lo tanto, cualquier componente principal con un valor propio de al menos uno, explica más de la varianza total que cualquiera de las variables originales. Así, una heurístico sencillo es seleccionar los componentes principales que tienen un valor propio de al menos uno. De la tabla 4.10 podemos observar que los únicos factores con valores propios mayores o iguales que uno son las componentes 1 y 2 que tienen valores propios de magnitud 6.816 y 2.76 respectivamente.
- Otro criterio originalmente propuesto por Cattell (1966) consiste en ordenar los valores propios en forma creciente, y encontrar dentro del espectro un punto de ruptura que divida a los valores propios en grandes y pequeños, y finalmente elegir los valores propios correspondientes a los valores "grandes", este método se puede obtener al graficar los valores propios λ_k vs k para obtener el llamado gráfico de sedimentación. Al inspeccionar el gráfico de sedimentación de izquierda a derecha, se busca el punto de de inflexión en el que los valores propios dejan de formar una pendiente (la gráfica se aplanan) y comienzan a describir una caída de poca inclinación. La figura 4.4 muestra el gráfico de sedimentación de los resultados obtenidos en nuestro análisis. Se puede observar que la pendiente pierde inclinación a partir del tercer valor propio (hacia su derecha). O lo que es lo mismo, el tercer valor propio no provoca pendiente respecto del cuarto (es decir, el tercer valor se encuentra en la misma planicie sin pendiente que el cuarto), por lo que debemos considerar que *sólo deben extraerse o retener los dos primeros factores y desechar del tercero en adelante.*

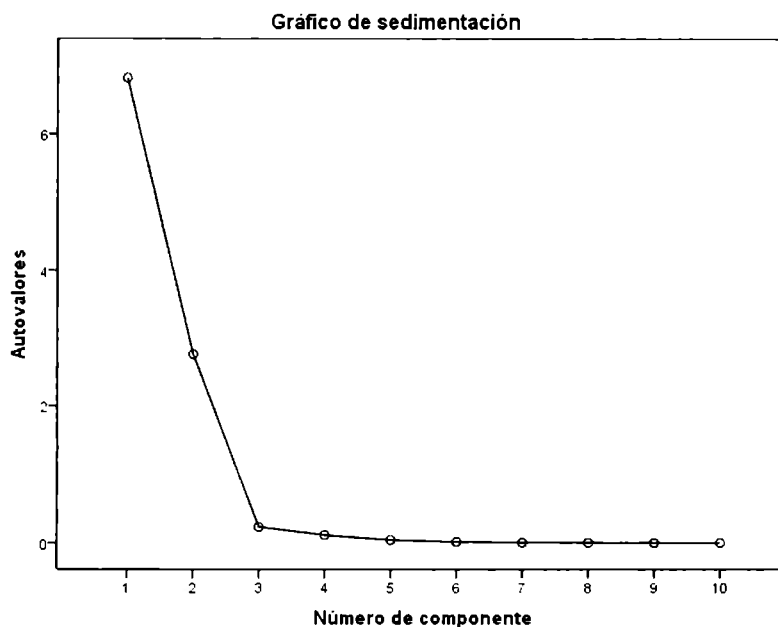


Figura 4.4: Gráfico de sedimentación.

4.2.5.4. Interpretación de los componentes principales.

Los principales aspectos a destacar que surgen de los resultados obtenidos por el ACP son los siguientes:

1. De la tabla 4.10 podemos observar que las cargas factoriales del primer componente principal llamado **nivel** son todas positivas y similares entre si, por lo que estos valores pueden ser entendidos como elementos de talla. Esto significa que frente a una perturbación aleatoria de digamos, $C_t^1 = 1$ (ver figura 4.5) las tasas de todos los bonos aumentan en una cantidades igual a las cargas factoriales, y como estas son similares se produce un desplazamiento casi paralelo de la curva de rendimiento, es por ello que este primer factor se asocia con desplazamientos paralelos.

Cabe destacar que en nuestro caso este factor explica un 68.2%, de la variabilidad total, siendo claramente el más importante. Se debe señalar además que el desplazamiento inducido por este primer componente no es exactamente paralelo⁷ ya que presenta una pequeña curvatura, siendo menor a largo plazo después a corto plazo,

⁷Para ello todas las cargas factoriales deberían ser iguales.

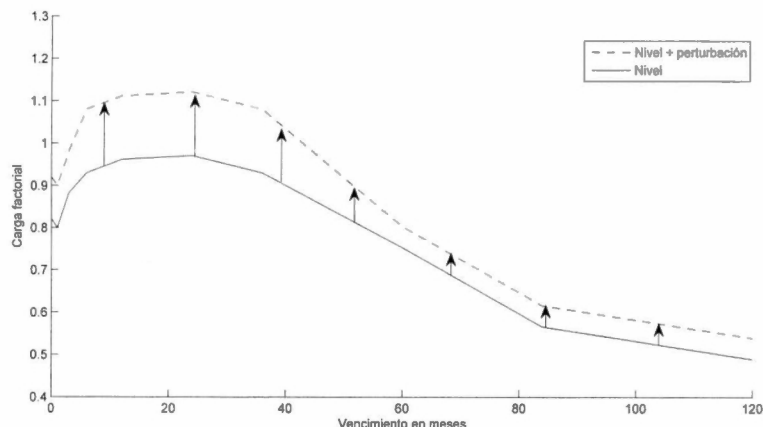


Figura 4.5: Primer componente principal.

y mayor a mediano plazo. Para el caso de México en el periodo de estudio, el primer componente principal⁸ o nivel tiene un comportamiento similar a la serie de tiempo de los rendimientos a 2 años, como se puede apreciar en la tabla 4.10 con la carga factorial entre ambas variables. Ang y Piazzesi (2003), Diebold, Piazzesi y Rudebusch (2005) utilizan como proxy del factor nivel el promedio de los rendimientos a corto, mediano y largo plazo, es decir, el promedio de los rendimientos a 1 día, 728 y 3600 días en nuestro caso. La correlación entre el 1er componente y éste promedio es de 0.9315. Sin embargo, hemos encontrado que en nuestro caso particular el promedio de todas las series a lo largo del espectro temporal es la mejor aproximación para el primer componente principal, hecho que se soporta por su alta correlación 0.992.

La figura 4.6 muestra la relación entre las series a lo largo del periodo de muestreo, esto sugiere que el promedio de los rendimientos tiende a marcar la pauta del **nivel** de la curva de rendimiento.

- Con respecto al segundo componente se observa el siguiente patrón, para los vencimientos a corto plazo prevalecen los valores negativos y para los vencimientos a largo plazo las cargas factoriales son positivas. Esto implica que si se perturbara el sistema con $C_t^2 = 1$, las tasas de los bonos a corto plazo decrecerán y las tasas de los bonos a largo plazo aumentarán, lo que podría interpretarse como un aumento en la

⁸Los coeficientes del primer componente fueron re-escalados por el valor característico λ_1 .

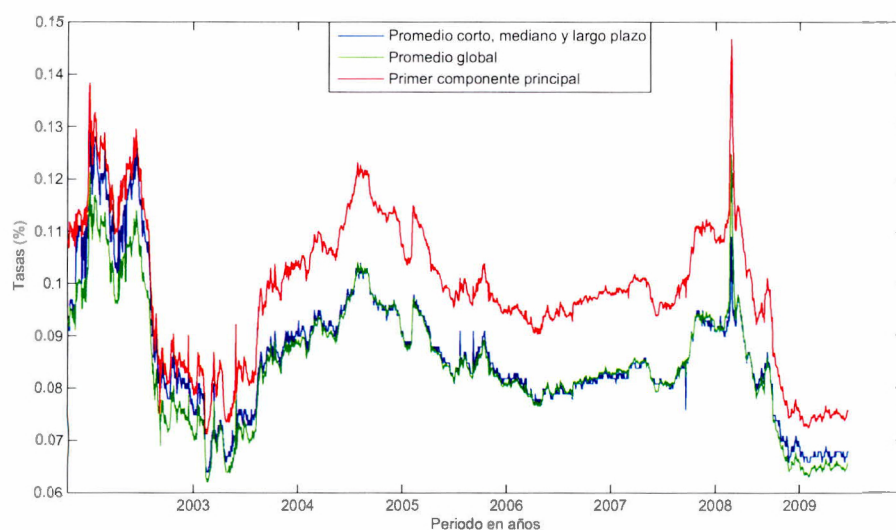


Figura 4.6: Comportamiento del 1er componente y el promedios total de las series de rendimientos.

pendiente de la curva de rendimiento. Este factor contribuye a explicar la variabilidad de las tasas en un 27.6% y gráficamente se manifiesta como si la componente girara en torno a un punto al perturbarse ante un choque.

Para el caso Mexicano, la pendiente o segundo componente principal puede ser considerado como un proxy de la diferencia entre los rendimientos a 10 años y el rendimiento a un 1 día, Ang y Piazzesi (2003), Diebold, Piazzesi y Rudebusch (2005), han utilizado esta aproximación. La correlación entre la diferencia (10 años menos 1 año) y la segunda componente es de 0.986 lo cual soporta la interpretación del segundo componente principal como factor de pendiente. En la figura 4.8 se puede observar que los movimientos de baja frecuencia muestran una reducción gradual de la pendiente de la curva de rendimiento que puede ser explicada por una reducción de la inflación y las expectativas de inflación durante el periodo de muestra.

3. El tercer componente, contribuye a la variabilidad total en un 2.3% y se le asocia a cambios en la concavidad de la curva, nuestros resultados permiten inferir dicha asociación, ya que los coeficientes de la tercera componente son positivos a corto plazo, negativos a mediano plazo y positivos a largo plazo. Si el sistema se perturba

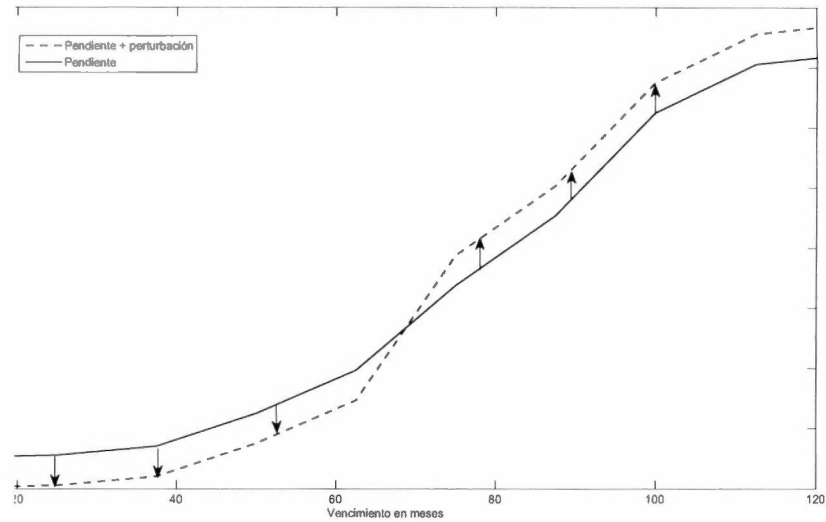


Figura 4.7: Segunda componente principal.

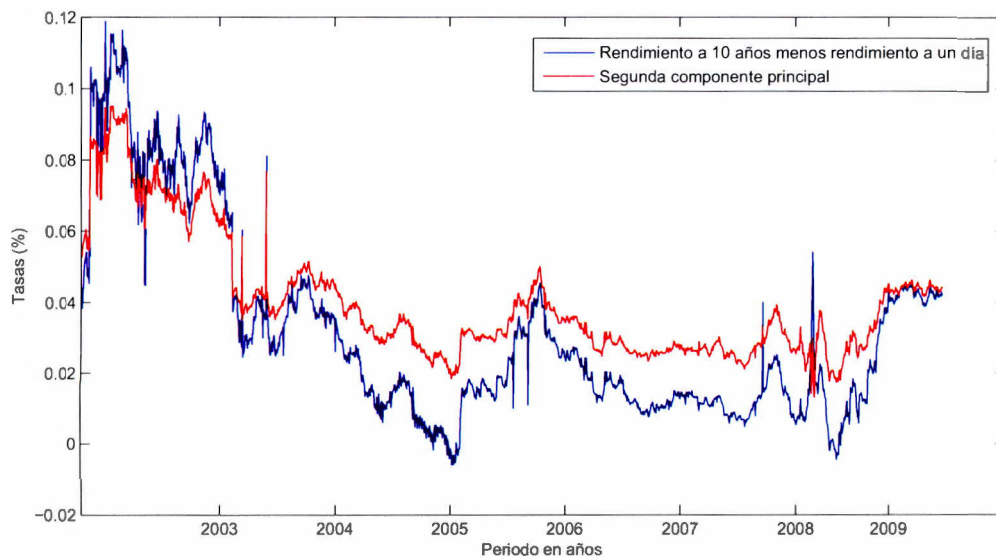


Figura 4.8: Comportamiento del 2do componente y el spread rendimiento a 10 años menos rendimiento a 1 día.

con $C_t^3 = 1$ las tasas a corto plazo aumentan, a mediano plazo decrecen y a largo plazo se incrementan lo que origina que la curvatura se incremente como se puede observar en la figura 4.9.

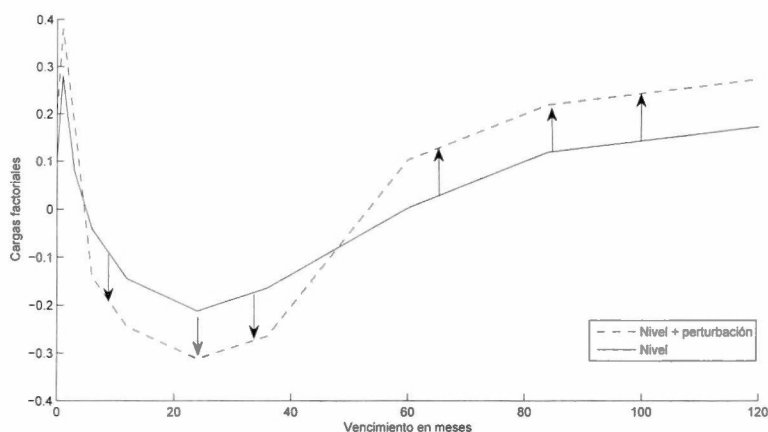


Figura 4.9: Tercer componente principal.

Como se ha señalado, el primer factor es casi plano, el segundo tiene signos opuestos en ambos extremos del espectro de madurez, y el tercero, tiene el mismo signo en ambos extremos, pero tiene un signo contrario en el centro. Esta observación nos lleva a considerar el número de cambios de signo de cada componente principal o vector propio. Si los primeros tres factores tienen cero, uno y dos cambios de signo, diremos que estamos en condiciones de observar el nivel, la pendiente y la curvatura respectivamente.

Los resultados anteriores también permiten explicar en parte un hecho conocido en los mercados, que afirma que cuando las tasas suben la curva de rendimientos se hace más plana (y en algunos casos se invierte) y cuando las mismas bajan, la curva tiende a adquirir mayor pendiente. Nótese, que el primer componente (el más importante) las cargas factoriales para los bonos a corto plazo son mayores que los correspondientes a 5, 7 y 10 años. En consecuencia cuando las tasas bajan, los bonos a corto plazo caen más que los bonos a largo plazo, inclinando más la curva. En cambio cuando suben las tasas, los bonos a corto plazo crecen más que los bonos a largo plazo aplanando la misma.

En síntesis, los tres primeros componentes explican un 98.1% de la varianza total de la curva de rendimiento de los CETES, por lo cual no parece relevante tomar en cuenta los restantes componentes para efectos de explicar los cambios en la curva de rendimiento,

además que la variabilidad residual se reparte entre los restantes componentes.

Estos resultados apoyan el hecho que los desplazamientos de la curva se interpreten como un desplazamiento de nivel que es casi como un corrimiento paralelo (1er comp.), un cambio de pendiente (2do comp.) y un cambio de curvatura (3er comp.). Por la forma como se definen los componentes principales, cada uno de estos movimientos son independientes el uno del otro, su representación gráfica se muestra en la siguiente figura.

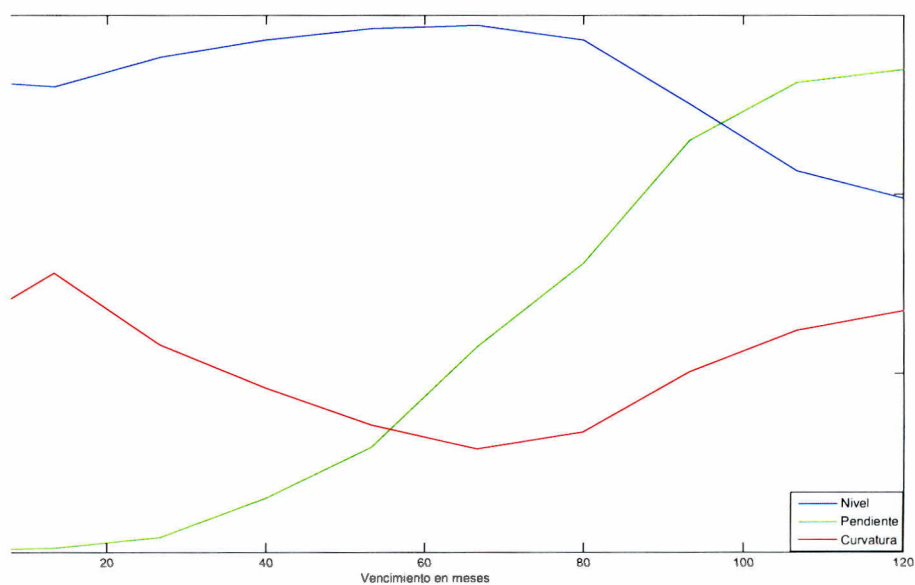


Figura 4.10: Vectores característicos de los tres primeros componentes principales.

Prueba de especificación no paramétrica para modelos de tiempo continuo.

5.1. Introducción.

LA estructura temporal de las tasas de interés ha recibido atención constante durante las últimas décadas, tanto por profesionales como por investigadores. Esto se debe a las numerosas aplicaciones que tiene la dinámica de la estructura temporal. Desde el punto de vista de la Economía Financiera, permite valorar múltiples activos y diseñar estrategias de inversión o de cobertura. Un enfoque común en la modelación de la estructura temporal de tasas de interés y de los precios de derivados de tasas de interés es la de expresar las tasas de interés en términos de uno o más factores aleatorios, los cuales siguen procesos estocásticos de tiempo continuo.

Los modelos de un factor son muy populares debido a su relativa sencillez matemática. En un modelo de un factor esencialmente se supone que la estructura temporal se ve afectada por una única fuente de incertidumbre que generalmente se toma como la tasa corta y todas las otras tasas de interés son funciones de ésta. Este es el caso, por ejemplo, de Merton (1973), Vasicek (1971), Dothan (1978), Marsh y Rosenfeld (1983), Cox, Ingersoll y Ross (1985), Hull y White (1990), Brown y Schaefer (1991), Aït-Sahalia(1996), Jiang y Knight (1991). Los modelos de un factor tienen la ventaja de ser parsimoniosos y puede asumir diferentes formas de la estructura de plazo, sin embargo pueden ser insuficientes

para describir la evolución completa de los movimientos de la estructura temporal de las tasas de interés. Una serie de estudios desarrollados por Dybvig (1989) y Steeley (1991) han concluido que la variabilidad de las tasas a diferentes vencimientos se puede explicar mejor y de una manera más real incorporando más de un factor estocástico. Esto ha llevado a varios autores a proponer modelos de la estructura temporal que incorporan dos o más factores estocásticos. Como se vio en el capítulo anterior, para el caso particular de México es necesario considerar dos o más factores para describir la dinámica de las tasas de interés.

Dentro de los modelos multi-factor se incluyen el de Brennan y Schwartz (1979), Schaefer y Schwartz (1984), Longstaff y Schwartz (1992), Hull y White (1990), Fong y Vasicek (1991), Heath, Jarrow y Morton (1992), Duffie y Kan (1995). El modelo descrito por Brennan y Schwartz (1979), por ejemplo, asume que ambas la tasa corta y la tasa de largo plazo son los motores de la dinámica de la estructura temporal, mientras que el presentado por Fong y Vasicek (1991) tiene la tasa corta y la volatilidad de ésta como factores principales.

Estos modelos de dos factores generalmente se excluyen mutuamente y generan precios significativamente diferentes cuando se utilizan para estimar el precio de derivados de tasas de interés. Así, se plantea la pregunta de cuál de los modelos se debe utilizar en la práctica. En ausencia de una teoría que nos diga que especificación particular debe tener un modelo de la estructura de plazos, la pregunta debe resolverse llevando los modelos a los datos, y en consecuencia, es deseable poder probar la especificación de los modelos de la estructura de plazos sin hacer uso de observaciones que provengan de una colección de derivados de tasas de interés. Aït-Sahalia (1996) provee un primer paso en esta dirección. Su prueba se basa en la siguiente idea: las propiedades de un modelo de la estructura de plazos están determinadas enteramente por un proceso de difusión, el cual está caracterizado por sus primeros dos momentos de tiempo continuo, la tendencia y la difusión. Cada modelo paramétrico de la estructura de plazos posee una cierta función de densidad que se caracteriza por las funciones de tendencia y difusión.

La prueba estadística de Aït-Sahalia (1996) se basa en una comparación de la función de densidad obtenida a partir del modelo paramétrico de la estructura temporal y un

estimador no paramétrico de la función de densidad obtenida a partir de los datos, la cual es válida incluso si el modelo paramétrico de la estructura temporal está mal especificado.

Sin embargo, hay que señalar que existen las siguientes limitaciones en la prueba: En primer lugar, se aplica sólo a modelos de un sólo factor de la estructura de plazos. En segundo lugar, el supuesto más importante en la prueba es que los datos deben estar subsuavizados cuando se construye el estimador no paramétrico de la función de densidad, es decir, el estimador no paramétrico es sub-óptimo. Por lo tanto, no está claro lo que le sucede a la prueba cuando los datos están sobresuavizados. En tercer lugar, la prueba no considera los efectos del sesgo inherente en la estimación no paramétrica de la función de densidad. En este capítulo, vamos a ampliar la prueba de Yacine Aït-Sahalia(1996) a lo largo de las tres limitaciones antes mencionadas.

5.2. El modelo y la hipótesis nula.

Para un espacio de probabilidad completo (Ω, \mathcal{F}, P) y filtración aumentada $\{\mathcal{F}_t : t \geq 0\}$ generada por un movimiento Browniano estándar W en \mathbb{R}^d , un modelo de tiempo continuo típicamente depende de un proceso de difusión estacionario X que toma valores en algún subconjunto abierto D de \mathbb{R}^d , con dinámica representada por una ecuación diferencial estocástica de Ito,

$$dX_t = \mu(X_t, \beta)dt + \sigma(X_t, \beta)dW_t \quad (5.1)$$

donde para cualquier β en un subconjunto acotado $\Theta \subset \mathbb{R}^d$, $\mu(\cdot, \beta) \in \mathbb{R}^d$ y $\sigma(\cdot, \beta) \in \mathbb{R}^{d \times d}$ son las funciones de tendencia y difusión respectivamente. La distribución del proceso se caracteriza en su totalidad por la tendencia y la difusión. Por ejemplo, para un proceso de difusión estacionario de una dimensión, la función de densidad marginal puede ser escrita de la siguiente manera

$$\pi(x, \beta) = \frac{\eta(\beta)}{\sigma^2(x, \beta)} \int_{x_0}^x \exp\left\{\frac{2\mu(u, \beta)}{\sigma^2(u, \beta)}\right\} du \quad (5.2)$$

donde el proceso se distribuye sobre \mathbb{R} y $\eta(\beta)$ es una constante de normalización es una constante de normalización que asegura que la densidad integra a uno.

Generalmente, para cualquier $\beta \in \Theta$ usamos $\pi(x, \beta)$ para expresar la función de densidad marginal implícita en el modelo paramétrico y $\pi(x)$ para expresar la función de densidad marginal verdadera. La hipótesis nula y alternativa que nos interesa son,

$$H_0 : \text{existe } \beta_0 \in \Theta \text{ tal que } \pi(x, \beta_0) = \pi(x)$$

$$H_1 : \pi(\cdot, \beta) \neq \pi(\cdot) \text{ para cualquier } \beta \in \Theta$$

Como en Aït-Sahalia(1996), nuestro estadístico de prueba se basa en la integral ponderada de la diferencia al cuadrado entre $\pi(x)$ y $\pi(x, \beta_0)$,

$$I = \int (\pi(x) - \pi(x, \beta))^2 \pi(x) dx \quad (5.3)$$

Podemos utilizar la medida I como un indicador de la especificación incorrecta del modelo, ya que $I \geq 0$ y $I = 0$ si y sólo si la función de densidad marginal implicada por el modelo se especifica correctamente. Para obtener un estimador consistente del parámetro β_0 , se utiliza el método no paramétrico *Delta de funcionales kernel* de Yacine Aït-Sahalia(1992) como sigue,

$$\hat{\beta}_n = \min_{\beta \in \Theta} L_n(\beta) \quad (5.4)$$

donde

$$L_n(\beta) = \frac{1}{n} \sum_{i=1}^n (\pi(x_i, \beta) - \hat{\pi}(x_i))^2 \quad (5.5)$$

y $\hat{\pi}(x)$ representa el estimador kernel de la función de densidad marginal $\pi(x)$, i.e.,

$$\hat{\pi}(x) = \frac{1}{n h^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \quad (5.6)$$

donde $K(\cdot)$ es una función kernel y $h = h_n$ es un parámetro de suavizado que no depende de los datos. Por lo tanto, bajo H_0 el estimador por el método Delta de $\pi(x, \beta)$ es $\pi(x, \hat{\beta}_n)$ y la función de densidad marginal verdadera $\pi(x)$, puede ser estimada consistentemente por el estimador kernel $\hat{\pi}(x)$, independientemente de si el modelo paramétrico

está especificado correctamente o no.

Si los estimadores $\hat{\pi}(x)$ y $\pi(x, \hat{\beta}_n)$ de $\pi(x)$ y $\pi(x, \beta)$ respectivamente se sustituyen en la definición de I , producen el siguiente estimador de I :

$$I_n = \int (\hat{\pi}(x) - \pi(x, \hat{\beta}_n))^2 \hat{\pi}(x) dx \quad (5.7)$$

5.3. La distribución límite de I_n bajo H_0 .

Los siguientes supuestos se utilizan para obtener la distribución límite de I_n .

Supuesto 1. Para cualquier $\beta \in \Theta$, $\sigma(\cdot, \beta)$ es localmente acotada y Borel medible.

Supuesto 2. Sea q una medida de probabilidad inducida sobre \mathbb{R}^d por X_0 , entonces $\int_{\mathbb{R}^d} \lambda \phi dq = 0$ para cualquier función continua y acotada $\phi(x)$ en \mathbb{R}^d , donde λ es el generador infinitesimal generado por el proceso de difusión $\{X_t, t \geq 0\}$.

Supuesto 3. La función kernel $K(\cdot)$ es una función simétrica y acotada en \mathbb{R}^d que satisface:

$$\int |K(u)| du < \infty, \quad \|u\|^d |K(u)| \rightarrow 0 \text{ cuando } \|u\| \rightarrow \infty,$$

$$\int u_i K(u) du = 0, \quad \int u_i u_j K(u) du = 2k \delta_{ij}, \text{ para } 1 \leq i, j \leq d, \quad k \in \mathbb{R}^+$$

Supuesto 4. La función de densidad $\pi(\cdot)$ y su derivada parcial de segundo orden son acotadas y uniformemente continuas en \mathbb{R}^d

Supuesto 5. El parámetro de suavizado $h = h_n$ satisface $h \rightarrow 0$, $nh^d \rightarrow \infty$ cuando $n \rightarrow \infty$.

Supuesto 6. La función de densidad paramétrica $\pi(u, \beta)$ y su derivada parcial de segundo orden con respecto a β son uniformemente acotadas y uniformemente continuas. $\pi(u, \beta)$ y su derivada parcial de segundo orden con respecto a x son acotadas y uniformemente continuas en \mathbb{R}^d . Además, $\int |D' \pi(u, \beta)| dx < \infty$, donde, $D' \pi(u, \beta)$ es un vector de orden

$p \times 1$ formado por las derivadas parciales de primer orden de la función $\pi(u, \beta)$ con respecto a β .

Supuesto 7. La sucesión de datos observados $\{X_i, 1 \leq i \leq n\}$ es estrictamente estacionaria.

Supuesto 8. Existe $\beta_* \in \Theta$ tal que $\hat{\beta}_n \rightarrow \beta_*$ a.s y

$$\begin{aligned} \hat{\beta}_n &= \beta_* + \int_{-\infty}^{+\infty} \varphi_F(u) d(F_n(u) - F(u)) + o_p\left(n^{-\frac{1}{2}}\right) \\ &= \beta_* + \frac{1}{n} \sum_{i=1}^n (\varphi_F(X_i) - E\varphi_F(X_i)) + o_p\left(n^{-\frac{1}{2}}\right) \end{aligned} \quad (5.8)$$

donde $F(\cdot)$ y $\varphi_F(\cdot)$ son la función de densidad acumulada y la derivada de $F(\cdot)$ respectivamente, asociadas con la función de densidad marginal desconocida $\pi(\cdot)$.

Supuesto 9. El espacio de parámetros $\Theta \subset \mathbb{R}^d$ es compacto y $E[(\partial\pi(u, \beta_0)/\partial\beta)(\partial\pi(u, \beta_0)/\partial\beta)']$ es de rango completo.

- El supuesto 1 garantiza la existencia y unicidad de una solución para la ecuación diferencial estocástica. Dado que nuestro modelo es una ecuación diferencial estocástica homogénea en el tiempo y del tipo Markoviano, el supuesto 1 también es una condición suficiente para una solución no explosiva.
- El supuesto 2 asegura que el proceso solución es estacionario.
- Los supuestos 3, 4 y 5 se utilizan con el fin de obtener la distribución límite del error cuadrado integrable del estimador de la densidad kernel.
- El papel de los supuestos 6 y 8 es examinar el efecto que tiene la estimación de $\pi(u, \beta_0)$ por $\tilde{\pi}(x)$ sobre la distribución límite de I_n . En particular, la existencia del último término $o_P(1)$ en el lado derecho de la ecuación (5.8), se garantiza asumiendo que la derivada $\varphi_F(\cdot)$ es una función cadlag acotada.
- La función del supuesto 7 es restringir la cantidad de dependencia permitida en las observaciones discretas con el fin de que podamos utilizar la teoría asintótica.

- El supuesto 9 garantiza que el término lineal en la expansión de Taylor de un funcional es no degenerado. Si el término lineal se degenera entonces la distribución asintótica vendría dada por un término de mayor orden en la expansión de Taylor.

De la ecuación (5.7) podemos escribir la diferencia de cuadrado integrable entre $\hat{\pi}(x)$ y $\tilde{\pi}(x)$ de la siguiente manera,

$$\begin{aligned} I_n &= \int (\hat{\pi}(x) - \pi(x))^2 \hat{\pi}(x) dx + \int (\tilde{\pi}(x) - \pi(x))^2 \hat{\pi}(x) dx \\ &- 2 \int (\hat{\pi}(x) - \pi(x))(\tilde{\pi}(x) - \pi(x))\hat{\pi}(x) dx = I_{1n} + I_{2n} - 2I_{3n} \end{aligned} \quad (5.9)$$

Teorema 1. Sea $c(n) = (nh^d)^{-1} \int K^2(x) dx \int \pi^2(x) dx + \int (E\hat{\pi}(x) - \pi(x))^2 \pi(x) dx$, $\nabla \pi(x) = \sum_{i=1}^d \partial^2 \pi(x) / \partial x_i^2$ y definamos

$$d(n) = \begin{cases} n^{1/2} h^{-2} & \text{si } nh^{d+4} \rightarrow \infty \\ nh^{d/2} & \text{si } nh^{d+4} \rightarrow 0 \end{cases}$$

entonces bajo las condiciones establecidas sobre $K(\cdot)$ y $\pi(\cdot)$ y asumiendo que $nh^{2d} \rightarrow \infty$, cuando $n \rightarrow \infty$, tenemos

a)

$$\begin{aligned} d(n)(I_{1n} - c(n)) &= 2d(n)(n^2 h^{2d})^{-1} \sum_{1 \leq i \leq j \leq n} H_n(X_i, X_j) \\ &+ 2d(n) \int (\hat{\pi}(x) - E\hat{\pi}(x))(E\hat{\pi}(x) - \pi(x))\pi(x) dx + O_p(d(n)n^{-3/2}h^{-3d/2}) \end{aligned}$$

donde

$$H_n(X_i, X_j) = \int \left[K\left(\frac{x-X_i}{h}\right) - EK\left(\frac{x-X_i}{h}\right) \right] \left[K\left(\frac{x-X_j}{h}\right) - EK\left(\frac{x-X_j}{h}\right) \right] \pi(x) dx$$

b)

$$\begin{aligned} &n h^{d/2} \left[(n^2 h^{2d})^{-1} \sum_{1 \leq i \leq j \leq n} H_n(X_i, X_j) \right] \xrightarrow{d} \sqrt{2} \sigma_2 N \\ &n^{1/2} h^{-2} \int [\hat{\pi}(x)]^2 \pi^2(x) dx - \left\{ \int [\nabla^2 \pi(x)] \pi^2(x) dx \right\}^2 \end{aligned}$$

donde

$$\begin{aligned}\sigma_1^2 &= \int [\nabla^2 \pi(x)]^2 \pi^2(x) dx - \left\{ \int [\nabla^2 \pi(x)] \pi^2(x) dx \right\}^2 \\ \sigma_2^2 &= \int \pi^4(x) dx \left\{ \int \left[\int K(u) K(u+v) du \right]^2 dv \right\}\end{aligned}$$

c)

$$d(n)(I_{1n} - c(n)) \xrightarrow{d} \begin{cases} 2k\sigma_1 N & \text{si } nh^{d+4} \rightarrow \infty \\ \sqrt{2}\sigma_2 N & \text{si } nh^{d+4} \rightarrow 0 \end{cases}$$

El teorema anterior implica que la distribución límite del error cuadrado integrable I_{1n} de $\hat{\pi}(x)$, depende de la cantidad de suavizado que se aplica a los datos. Las distribuciones límite del teorema 1 pueden ser usadas para construir la prueba de la hipótesis nula $\pi(x) = \pi_0(x)$ contra la alternativa $\pi(x) \neq \pi_0(x)$, donde $\pi_0(x)$ es una función de densidad completamente desconocida. Además, cabe señalar que, para que el teorema 1 se cumpla, el parámetro de suavizado h debe satisfacer $nh^{2d} \rightarrow \infty$ o $nh^{d+4} \rightarrow 0$ (datos subsuavizados) o $nh^{d+4} \rightarrow \infty$ (datos sobresuavizados). Así que d debe satisfacer $d \leq 3$ para datos subsuavizados. Sin embargo, cuando la hipótesis nula es compuesta, es decir, cuando hay un número finito de parámetros desconocidos, los dos últimos términos en el lado derecho de la ecuación (5.9) debe tomarse en consideración. Para este propósito, por la expansión de Taylor y el supuesto 8, bajo H_0 tenemos

$$\hat{\pi}(x) - \pi(x, \beta_0) = D\pi(x, \beta_0) \frac{1}{n} \sum (\varphi_F(X_i) - E\varphi_F(X_i)) + o_p(n^{-1/2})$$

entonces expresamos I_{3n} de la siguiente manera,

$$I_{3n} \equiv J_{1n}(\beta_0) + J_{2n}(\beta_0)$$

donde

$$J_{1n}(\beta_0) = \int [\hat{\pi}(x) - E\hat{\pi}(x)][\hat{\pi}(x) - \pi(x, \beta_0)]\hat{\pi}(x) dx \quad (5.10)$$

y

$$J_{2n}(\beta_0) = \int [E\hat{\pi}(x) - \pi(x)][\hat{\pi}(x) - \pi(x, \beta_0)]\hat{\pi}(x)dx \quad (5.11)$$

El siguiente lema resume el comportamiento de $J_{1n}(\beta_0)$ y $J_{2n}(\beta_0)$, y por lo tanto de I_{3n} .

Lema 2. *Bajo los supuestos 1-9, para cualquier $0 < \delta < 1$ se tiene*

a)

$$J_{1n}(\beta_0) = O_p(n^{-\delta}) \text{ y además } n^{1/2}h^{-2}J_{2n}(\beta_0) \xrightarrow{d} k\sigma_3N$$

donde

$$\sigma_3^2 = \left\{ \int D_{\beta_0}\pi(x, \beta_0)\nabla^2\pi(x)dx \right\} V \left\{ \int D_{\beta_0}\pi(x, \beta_0)\nabla^2\pi(x)dx \right\}$$

$$V = Var(\varphi_F(x_1)) + \sum_{k=1}^{\infty} Cov(\varphi_F(X_1), \varphi_F(X_{1+k}))$$

consecuentemente

$$I_{3n} = J_{2n}(\beta_0) + O_p(n^\delta) \text{ y } n^{1/2}h^{-2}I_{3n} \xrightarrow{d} k\sigma_3N$$

b) *Si la hipótesis nula se satisface, $I_{2n} = O_p(n^{-1})$ y también $n^{1/2}h^{-2}I_{3n} \xrightarrow{d} k\sigma_3N$*

Supongamos que la hipótesis nula se cumple, entonces de la ecuación (5.11) y el lema anterior se tiene la siguiente expresión,

$$I_n = I_{1n} - 2J_{2n}(\beta_0) + O_p(n^{-\delta}) \quad (5.12)$$

la ecuación anterior establece que la estimación paramétrica de $\pi(x, \beta_0)$ bajo los efectos de la hipótesis nula es la distribución límite de I_n sólo a través de $J_{2n}(\beta_0)$. Por lo tanto, si el primer término en el lado derecho de la ecuación (5.12) domina al segundo término asintóticamente que es el caso cuando los datos son subsuavizados, la estimación paramétrica de $\pi(x, \beta_0)$ bajo la hipótesis nula no afecta la distribución límite de I_n como lo demostró Yacine Aït-Sahalia (1996) para $d = 1$. Sin embargo, si los datos son sobresuavizados, la

distribución límite de I_n se verá afectada por los estimadores no paramétricos, como se demuestra en el siguiente teorema.

Teorema 2. *Bajo los supuestos del lema, si la hipótesis nula se satisface y $\frac{d + \delta}{2(d + 4)} \leq \delta < 1$, entonces se tiene*

$$d(n)(I_n - c(n)) \xrightarrow{d} \begin{cases} 2k\sigma_4 N & \text{si } nh^{d+4} \rightarrow \infty \\ \sqrt{2}\sigma_2 N & \text{si } nh^{d+4} \rightarrow 0 \end{cases}$$

donde $c(n), d(n), \sigma_1^2, \sigma_2^2$ son los mismos que en el teorema 1, $\sigma_4^2 = V_1 + 2V_2$ y:

$$V_1 = \text{Var}(\nabla^2 \pi(X_1)) - 2 \int D' \pi(y, \beta) \nabla^2 \pi(y) dy + \int D' \pi(y, \beta) \nabla^2 \pi(y) dy \times \text{Var}(\varphi_F(X_1)) \times \int D \pi(y, \beta) \nabla^2 \pi(y) dy \quad (5.13)$$

$$V_2 = \sum_{k=2}^{\infty} \text{Cov}(\nabla^2 \pi(X_1), \nabla^2 \pi(X_k)) + \left(\int D' \pi(y, \beta) \nabla^2 \pi(x) dx \right) \times \sum_{k=2}^{\infty} \text{Cov}(\varphi_F(X_1), \varphi_F(X_k)) \times \left(\int D \pi(y, \beta) \nabla^2 \pi(y) dy \right) \quad (5.14)$$

$$- \int D' \pi(y, \beta) \nabla^2 \pi(y) dy \times \sum_{k=2}^{\infty} \text{Cov}(\nabla^2 \pi(X_1), \varphi_F(X_k)) - \int D' \pi(y, \beta) \nabla^2 \pi(y) dy \times \sum_{k=2}^{\infty} \text{Cov}(\nabla^2 \pi(X_k), \varphi_F(X_1)) \quad (5.15)$$

Corolario 1. *Bajo los supuestos del lema, y si la hipótesis nula se satisface, tenemos*

a)

$$\text{Si } nh^{d+4} \rightarrow \infty, \text{ entonces } n^{1/2} h^{-2} [I_n - \int [E \hat{\pi}(x) - \pi(x, \beta_0)]^2 \hat{\pi}(x) dx] \xrightarrow{d} 2k\sigma_4 N$$

b)

$$\text{Si } nh^{d/2+4} \rightarrow 0, \text{ entonces } nh^{d/2} [I_n - \frac{1}{nh^d} \int K^2(u) du \int \pi^2(u) du] \xrightarrow{d} \sqrt{2}\sigma_2 N$$

5.4. Prueba estadística y sus distribuciones asintóticas.

Las pruebas estadísticas definidas a continuación se basan en los estimadores de σ_2^2 y $\sigma_4^2 \cdot \sigma_2^2$ pueden ser estimados consistentemente por $\frac{1}{n} \sum_{i=1}^n \hat{\pi}^3(X_i) \int [\int K(u)K(u+v)du]^2 dv$. Para obtener estimadores consistentes de σ_4^2 , definimos

$$\widehat{Var}(\nabla^2 \pi(X_1)) = \int (\nabla^2 \hat{\pi}(x))^2 \hat{\pi}(x) dx - \left\{ \int (\nabla^2 \hat{\pi}(x)) \hat{\pi}(x) dx \right\};$$

$$\begin{aligned} & \widehat{\sum_{k=1}^{\infty} Cov}(\varphi_F(X_1), \varphi_F(X_{k+1})) = \\ & \sum_{k=1}^{G_n} \frac{2}{1-k+G_n} \left\{ \frac{1}{n} \sum_{i=1}^{n-k} \varphi_{\hat{F}_n}(X_{i+k}) - \left(\frac{1}{n} \sum_{i=1}^n \varphi_{\hat{F}_n}(X_i) \right)^2 \right\}; \end{aligned}$$

$$\begin{aligned} & \widehat{\sum_{k=1}^{\infty} Cov}(\nabla^2 \pi(X_1), \nabla^2 \pi(X_{k+1})) = \\ & \sum_{k=1}^{G_n} \frac{2}{1-k+G_n} \left\{ \frac{1}{n} \sum_{i=1}^{n-k} \nabla^2 \hat{\pi}(X_i) \nabla^2 \hat{\pi}(X_{i+k}) - \left(\frac{1}{n} \sum_{i=1}^n \nabla^2 \hat{\pi}(X_i) \right)^2 \right\}; \end{aligned}$$

$$\begin{aligned} & \widehat{\sum_{k=1}^{\infty} Cov}(\nabla^2 \pi(X_1), \varphi_F(X_{k+1})) = \\ & \sum_{k=1}^{G_n} \frac{2}{1-k+G_n} \left\{ \frac{1}{n} \sum_{i=1}^{n-k} \nabla^2 \hat{\pi}(X_i) \varphi_{\hat{F}_n}(X_{i+k}) - \frac{1}{n^2} \sum_{i=1}^n \nabla^2 \hat{\pi}(X_i) \sum_{i=1}^n \varphi_{\hat{F}_n}(X_{i+k}) \right\}; \end{aligned}$$

$$\begin{aligned} & \widehat{\sum_{k=1}^{\infty} Cov}(\nabla^2 \pi(X_{1+k}), \varphi_F(X_1)) = \\ & \sum_{k=1}^{G_n} \frac{2}{1-k+G_n} \left\{ \frac{1}{n} \sum_{i=1}^{n-k} \nabla^2 \hat{\pi}(X_{i+k}) \varphi_{\hat{F}_n}(X_i) - \frac{1}{n^2} \sum_{i=1}^n \nabla^2 \hat{\pi}(X_{i+k}) \sum_{i=1}^n \varphi_{\hat{F}_n}(X_i) \right\}, \end{aligned}$$

donde G_n es un rezago truncado, elegido de tal que $\lim G_n = +\infty$ y $G_n = O(n^{1/3})$. Este es un estimador de la densidad espectral en cero ver (Yacine Aït-Sahalia(1992)). Se pueden hallar estimadores consistentes de V_1 y V_2 sustituyendo los estimadores anteriores en las ecuaciones (5.13) y (5.14). Usamos los resultados obtenidos en el corolario para

generar pruebas estadísticas para la hipótesis $H_0 : \pi(x) = \pi(x, \beta_0)$ contra la alternativa general $H_1 : \pi(x) \neq \pi(x, \beta)$ para todo $\beta \in \Theta$.

5.5. Extensión del estadístico de prueba de Yacine Aït-Sahalia.

Ahora definimos dos estadísticos de prueba de la siguiente manera:

$$T_{1n} = \frac{n^{1/2}h^{-2}\{I_n - \int [K_h * \tilde{\pi}(x) - \tilde{\pi}(x)]^2 \tilde{\pi}(x) dx\}}{2k\hat{\sigma}_4} \quad (5.16)$$

$$T_{2n} = \frac{h^{-d/2}[nh^d I_n - \int K^2(z) dz \int \pi^2(z) dz]}{\sqrt{2}\hat{\sigma}_2} \quad (5.17)$$

donde $K_h * \tilde{\pi}(x) = h^{-d} \int K\left(\frac{x-u}{h}\right) \tilde{\pi}(u) du$

Teorema 3. *Bajo los supuestos 1 a 9, si H_0 se satisface y $\frac{d+\delta}{2(d+4)} \leq \delta < 1$ entonces*

(a)

$$\text{Si } nh^{d/2+4} \rightarrow \infty \text{ entonces } T_{1n} \xrightarrow{d} N$$

(b)

$$\text{Si } nh^{d/2+4} \rightarrow 0 \text{ entonces } T_{2n} \xrightarrow{d} N$$

El teorema 3 nos dice, que se pueden construir dos diferentes pruebas a un determinado nivel de significancia α para probar la hipótesis H_0 contra H_1 correspondientes a diferentes cantidades de suavizado aplicado a los datos. Además, cabe señalar que el estadístico T_{2n} es una extensión de la prueba de Yacine Aït-Sahalia(1996), y sólo se puede aplicar a modelos de dos factores, por el valor pequeño del parámetro de suavizado.

5.6. Estadístico de prueba con corrección del sesgo.

Como hemos señalado, el sesgo introducido por el estimador kernel de la función de densidad, tiene una influencia significativa en la prueba de H_0 si se utiliza la estimación de error cuadrado integrable. En particular, los efectos del sesgo se reflejan en las restricciones del parámetro de suavizado el cual puede tender a cero y no puede decrecer demasiado rápido, ni tampoco demasiado lento cuando el tamaño de la muestra tiende a infinito. En función de las diversas restricciones en el parámetro de suavizado, se obtienen dos pruebas diferentes. Ahora, presentamos un estadístico de prueba que no sólo se aplica a cualquier proceso de difusión de dimensión finita, sino que también no depende de la cantidad de suavizado que se aplica a los datos. Bajo H_0 , el valor esperado del estimador no paramétrico $\hat{\pi}(x)$ es $E[\hat{\pi}(x)] = K_h * \pi(x, \beta_0) = \int h^{-d} K(\frac{x-u}{h}) \pi(u, \beta_0) du$. Por lo tanto, para eliminar los efectos del sesgo, se puede construir el estadístico de prueba de sesgo corregido sobre la base de la integral ponderada de la diferencia al cuadrado entre el estimador no paramétrico de la función de densidad marginal implicada en los datos y el estimador kernel-suavizado de la función de densidad marginal implicada por el modelo paramétrico:

$$J_n = \int [\hat{\pi}(x) - K_h * \tilde{\pi}(x)]^2 dx$$

Podemos descomponer J_n como sigue

$$\begin{aligned} J_n &= \int [\hat{\pi}(x) - E\tilde{\pi}(x)]^2 \hat{\pi}(x) dx + \int [E\tilde{\pi}(x) - K_h * \tilde{\pi}(x)]^2 \hat{\pi}(x) dx \\ &\quad + 2 \int [\hat{\pi}(x) - E\tilde{\pi}(x)][E\tilde{\pi}(x) - K_h * \tilde{\pi}(x)] \hat{\pi}(x) dx \\ &= \int [\hat{\pi}(x) - E\tilde{\pi}(x)]^2 \hat{\pi}(x) dx + J_{11} + 2J_{12} \end{aligned}$$

Teorema 4. *Bajo los supuestos 1 a 9, si H_0 se cumple y $\frac{d+\delta}{2(d+\delta)} \leq \delta < 1$, entonces*

$$T_n = \frac{nh^{d/2} [J_n - \frac{1}{nh^d} \int K^2(z) dz \int \pi^2(z) dz]}{\sqrt{2\hat{\sigma}_2}} \xrightarrow{d} N \quad (5.18)$$

donde $\hat{\sigma}_2^2$ se define como antes.

La distribución asintótica de la prueba de sesgo corregido es la misma independiente-

mente de si los datos están subsuavizados o sobresuavizados. También vale la pena señalar que la prueba de sesgo corregido puede ser aplicada a la función de densidad marginal de cualquier proceso de difusión de dimensión finita.

5.7. Modelos paramétricos de dos factores.

Existe un consenso en que los modelos de un sólo factor de la estructura a plazos no pueden explicar con éxito varias de las principales características de los rendimientos de los bonos. La razón es que en un modelo de un factor, los rendimientos y los retornos de los bonos debe estar perfectamente correlacionados. Como hemos visto en el capítulo 4, varios autores sostiene que los rendimientos son impulsados por al menos dos factores de riesgo. Esto ha llevado a varios investigadores a proponer modelos de la estructura temporal que incorporan dos variables de estado o factores estocásticos. Estos modelos suelen ser mutuamente excluyentes y generan diferentes valores cuando se utilizan para hallar el precio de derivados de distintos valores o títulos. Dentro de los modelos de dos factores se encuentran el de Brennan y Schwartz (1979) y el de Schaefer Schwartz (1984) que basan su especificación en procesos estocásticos de dos tasas de interés. En esta sección, aplicamos nuestra prueba para probar la especificación de los dos modelos. Antes de embarcarse en la prueba de especificación del modelo de Brennan-Schwartz y del modelo de Schaefer-Schwartz, vamos a encontrar las funciones de densidad marginal de estos modelos.

5.7.1. Modelo de la estructura de plazos de Brennan-Schwartz y Schaefer-Schwartz.

5.7.1.1. Modelo de la estructura de plazos de Brennan - Schwartz

Brennan y Schwartz (1979) desarrollaron un modelo de arbitraje de la estructura temporal de las tasas de interés basado en el supuesto, que toda la estructura de plazos en cualquier punto en el tiempo se puede expresar en función de los rendimientos de los instrumentos a largo y a corto plazo. Si r denota la tasa de interés instantánea y l la tasa de

interés a largo plazo. Su modelo puede ser expresado de la siguiente manera,

$$dr_t = r[\alpha(\ln l_t - p - \ln r_t) + 1/2\sigma_r^2]dt + (\sigma_r \cdot r_t)dW_1 \quad (5.19)$$

$$dl_t = [l_t(k(\theta - \ln l_t) + 1/2\sigma_l^2)]dt + (\sigma_l \cdot l_t)dW_2 \quad (5.20)$$

donde p es la diferencia entre los niveles medios de $\ln l$ y $\ln r$; α es el coeficiente de ajuste de la velocidad con la que $\ln r$ regresará a $(\ln l - p)$; k es el coeficiente de ajuste de la velocidad con la que $\ln l$ regresará a su nivel promedio θ ; y $E(dW_1dW_2) = \delta$. Usando el lema de Ito, las ecuaciones anteriores pueden ser escritas como,

$$d \ln r_t = \alpha((\ln l_t - p) - \ln r_t)dt + \sigma_r dW_1 \quad (5.21)$$

$$d \ln l_t = k(\theta - \ln l_t)dt + \sigma_l dW_2 \quad (5.22)$$

Para poner en práctica nuestra prueba, tenemos que estimar los parámetros desconocidos bajo la hipótesis nula. Los procesos bivariados para $\ln r$ y $\ln l$ son estacionarios en forma conjunta si α y k son positivos. Además, la solución a las ecuaciones (5.21) y (5.22) es un proceso Gaussiano si y sólo si el valor inicial es constante o normalmente distribuido. En realidad la solución de las ecuaciones diferenciales estocásticas lineales (5.21) y (5.22) se puede expresar como

$$X_t = \mu + \exp(A(t - t_0)) \cdot (X_{t_0} - \mu) + \int \exp(A(t - s))B^{1/2}dW \quad (5.23)$$

donde

$$X_t = \begin{pmatrix} \ln r_t \\ \ln l_t \end{pmatrix}, \quad \mu = \begin{pmatrix} \theta - p \\ \theta \end{pmatrix}, \quad A = \begin{pmatrix} -\alpha & \alpha \\ 0 & -k \end{pmatrix}, \quad B = \begin{pmatrix} \sigma_r^2 & 0 \\ 0 & \sigma_l^2 \end{pmatrix},$$

$$dW = \begin{pmatrix} dW_1 \\ dW_2 \end{pmatrix}$$

Por otra parte, si α y K son positivos, entonces la solución sigue a un proceso Gaussiano estacionario. Para un proceso estocástico estacionario, la función de densidad marginal es

igual a la densidad de la observación inicial, por lo tanto, podemos obtener la función de densidad marginal estacionaria de este proceso una vez que tengamos la función de densidad de la variable aleatoria inicial. La densidad marginal puede ser obtenida de la ecuación (5.22) de la siguiente manera,

$$f(X) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp(-(X - \mu)'\Sigma^{-1}(X - \mu)) \quad (5.24)$$

donde

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad X = \begin{pmatrix} \ln r_t \\ \ln l_t \end{pmatrix}$$

$$\Sigma_{11} = \frac{\sigma_r^2}{2\alpha} + 2\sigma_r^2\sigma_l^2\delta\left(\frac{\alpha}{\alpha-k}\right)\left(\frac{1}{(\alpha+k)-\frac{1}{2\alpha}}\right) + \sigma_l^2\left(\frac{\alpha}{\alpha-k}\right)^2\left(\frac{1}{2k} - \frac{2}{(\alpha+k)} + \frac{1}{2\alpha}\right)$$

$$\Sigma_{12} = 2\sigma_r\sigma_l\delta\left(\frac{1}{\alpha+k}\right) + \sigma_l^2\left(\frac{\alpha}{\alpha-k}\right)\left(\frac{1}{2k} - \frac{1}{(\alpha+k)}\right)$$

$$\Sigma_{21} = \Sigma_{12}$$

$$\Sigma_{22} = \frac{\sigma_l^2}{2k}$$

5.7.1.2. Modelo de la estructura de plazos de Schaefer-Schwartz

Schaefer y Schwartz (1984) utilizan la misma información sobre las tasas de interés que Brennan y Schwartz (1979), pero expresan su modelo en términos de la tasa a largo plazo y la diferencia entre ésta y la tasa a corto plazo. Este cambio es en realidad una redefinición de las variables que hace posible obtener una solución analítica para el problema de valuación. Además, Schaefer y Schwartz asumen que la diferencia sigue un proceso de Ornstein-Uhlenbeck (una proceso con reversión a la media y función de difusión constante). En la literatura financiera, el proceso de Ornstein-Uhlenbeck se ha utilizado para modelar la tasa a corto plazo. Sin embargo, es probable que sea más razonable suponer que la diferencia más que la tasa corta sigue este tipo de procesos debido a que el modelo permite valores negativos. La forma específica del proceso estocástico asumido por el

modelo de Schaefer y Schwartz es,

$$dS_t = \beta(\alpha - S_t)dt + \gamma dW_1 \quad (5.25)$$

$$dl_t = \mu_2(S_t, l_t, t)dt + \sigma\sqrt{l_t} dW_2 \quad (5.26)$$

donde dW_1 y dW_2 son procesos de Wiener estándar con $E[dW_1] = E[dW_2] = 0$, $dW_1^2 = dW_2^2 = dt$. Al igual que en el modelo de Brennan-Schwartz, Schaefer y Schwartz asumen que la función de difusión de la tasa console depende de su nivel. La función de tendencia en la tasa console se queda como en su forma general. El proceso de Ornstein-Uhlenbeck, $dS_t = \beta(\alpha - S_t)dt + \gamma dW_1$, tiene una función de densidad de transición dada por $f(S_t = s, t | S_{t_0} = s_0, t_0) = \frac{1}{\sqrt{2\nu^2(t)}} \exp\left\{-\frac{(s-\alpha - (s_0-\alpha)e^{-\beta(t-t_0)})^2}{2\nu^2(t)}\right\}$, donde $\nu^2(t) = \frac{\sigma^2}{2\beta}[1 - e^{-2\beta(t-t_0)}]$.

Supongamos que el proceso spread muestra la propiedad de reversión a la media ($\beta > 0$), entonces cuando $t_0 \rightarrow -\infty$ o $(t-t_0) \rightarrow \infty$, la densidad marginal del proceso estocástico es invariante en el tiempo, i.e., el proceso spread es estacionario y la función de densidad marginal puede ser expresada de la siguiente manera:

$$\sqrt{\frac{\pi\sigma^2}{\beta}} \exp\left\{-\frac{\beta(s-\alpha)^2}{\sigma^2}\right\}$$

Los únicos procesos estacionarios con funciones de densidad de transición conocidas de manera explícita, son aquellos que tienen la especificación que su función de tendencia es lineal y la especificación que su función de difusión es una función cuadrática.

5.7.2. Pruebas empíricas y análisis de especificación de la estructura de plazos de los modelos de Brennan-Schwartz y Schaefer-Schwartz

Las series de tiempo que utilizamos en este análisis incluyen observaciones diarias de los CETES para los vencimientos a 1 día y 10 años. El período muestral analizado se extiende del 9 de julio de 2002 hasta el 7 de diciembre de 2009, lo que recoge una muestra de 1867 observaciones para cada plazo. Con base en las funciones de densidad estaciona-

rias derivada anteriormente, podemos estimar los parámetros desconocidos, reduciendo al mínimo el error cuadrático entre estimador de la densidad no-paramétrica y la densidad del modelo paramétrico. Los resultados de la estimación paramétrica de los dos modelos se presentan en la tabla 5.1 y la tabla 5.2.

Parámetro	α	k	θ	ρ	σ_r	σ_l	δ
Estimación	0.98725	0.23873	0.10191	-0.01533	0.4956	0.21096	0.24627

Tabla 5.1: Valores de los parámetros estimados en el modelo de Brennan y Schwartz

Parámetro	α	β	σ
Estimación	1.7612	1.50114	0.1796

Tabla 5.2: Valores de los parámetros estimados en el modelo de Schaefer y Schwartz

Los estimadores de $(\int K^2(z)dz)(\int \pi^2(x)dx)$ y $(\int(\int K(u)K(u+x)du)^2)(\int \pi^4(x)dx)$ son respectivamente:

$$\int K^2(z)dz \left(\frac{1}{n} \sum_{i=1}^n \hat{\pi}_0(\log(r_i), \log(l_i)) \right)$$

$$\left(\int (\int K(u)K(u+x)du)^2 dx \right) \left(\frac{1}{n} \sum_{i=1}^n \hat{\pi}_0^3(\log(r_i), \log(l_i)) \right)$$

Tomando un valor crítico de 6.32 para obtener una prueba al nivel de 0.05%. Para el kernel normal, las dos constantes kernel son:

$$\int K^2(z)dz = \begin{cases} \frac{1}{2\sqrt{\pi}} & \text{si } d = 1 \\ \frac{1}{4\pi} & \text{si } d = 2 \end{cases}, \quad \int (\int K(u)K(u+x)du)^2 dx = \begin{cases} \frac{1}{2\sqrt{2\pi}} & \text{si } d = 1 \\ \frac{1}{8\pi} & \text{si } d = 2 \end{cases}$$

Para estimar la densidad marginal, el parámetro de suavizado $h_n = h$, se elige de manera que $\lim_{n \rightarrow \infty} nh^2 = \infty$ y $\lim_{n \rightarrow \infty} nh^5 = 0$. La calidad de la estimación de la densidad depende principalmente de la elección del parámetro de suavizado más que del kernel. Seleccionamos $h = c_n n^{-1/5}$, donde $c_n = c(\ln(n))$. c se elige de tal manera que minimice el error cuadrado medio integrable del estimador. Los resultados de la prueba se muestran en la tabla 5.3.

Los modelos de Brennan-Schwartz y Schaefer-Schwartz se rechazan al nivel de significancia de 0.05 %. Las principales fuentes de rechazo de ambos modelos son:

- Primero, las difusiones constantes limitan las volatilidades a ser uniformemente crecientes.
- Segundo, si un modelo posee una especificación lineal en la tendencia (drift) y tiene difusión constante, como en el caso de Brennan-Schwartz y Schwartz-Schaefer, los parámetros que se encuentran en la función de tendencia y difusión pueden no ser homogéneos, es decir, cambios de régimen económico implica la posible falta de estacionariedad de las parámetros durante el proceso de modelado.

Brennan-Schwartz					
Función de tendencia	Función de difusión	Estadístico de prueba I T_{2n}	Estadístico de prueba II T_n	Valor crítico	Resultado
$r[\alpha(\ln l - p - \ln r) + 1/2\sigma_r^2]$	$\sigma_r \cdot r$	46.2792		1.645	Rechazar
$[l(k(\theta - \ln l))] + 1/2\sigma_l^2]$	$\sigma_l \cdot l$		21.48973	1.645	Rechazar
Schaefer-Schwartz					
$\beta(\alpha - S_t)$	$\frac{\gamma}{\sigma\sqrt{l}}$	75.07685		1.645	Rechazar
$\mu(s, l, t)$			123.97681	1.645	Rechazar

Tabla 5.3: Pruebas de especificación paramétrica de modelos de tiempo continuo de dos factores.

Conclusiones

Los resultados encontrados producto del análisis de componentes principales revelan que para analizar la dinámica de la curva de rendimiento, es posible simplificar el estudio a tan solo dos variables que se hallan latentes en los datos: el nivel y la pendiente de la curva de rendimiento.

El factor preponderante el cual explica un poco más del 68% de los cambios se halla asociado al nivel de la curva, y está determinado principalmente por la tasa de interés de menor plazo y por choques estructurales que desplazan paralelamente a la curva.

Por su parte la pendiente que explica el 27.6%, usualmente se calcula como la diferencia entre el rendimiento de mayor plazo y el de menor vencimiento. Los cambios en la pendiente, implican que factores macroeconómicos, tales como el nivel de precios de la economía, la actividad económica y las tasas de interés a corto plazo (instrumentos con vencimientos menores a un año) están detrás de ese movimiento, además que resume las expectativas de los inversionistas sobre las tasas de interés futuras.

Finalmente como parte del diseño, al aplicar en análisis de componentes principales a la muestra es posible concluir lo siguiente:

- Las cargas factoriales son mayores si se utiliza la matriz de covarianza en lugar de la matriz de correlación.
- Al estandarizar las variables, los resultados obtenidos con la matriz de correlación coinciden con los resultados generados con la matriz de varianza-covarianza.

- Las cargas son mayores si se utilizan los niveles (datos crudos) en lugar de los cambios.
- El ACP es sensible al tamaño de la muestra.
- Al correr el análisis con tasas equivalentes los resultados no difieren de los resultados obtenidos al estandarizar las variables.
- Al quitar la serie de un día se obtienen cargas más grandes que al incluirla, es decir, la estructura de correlación entre las variables de la muestra depende de la tasa a un día.

Para la muestra que va del 9 de julio de 2002 hasta el 7 de diciembre de 2009 de los rendimientos diarios de los CETES correspondientes a los plazos de 1 día, 1, 3 y 6 meses, y 1, 2, 3, 5, 7 y 10 años, encontramos que el nivel puede ser aproximado por el promedio de las series de los rendimientos que conforman la muestra, mientras que la pendiente puede ser aproximada por el spread entre los rendimientos a 10 años y los rendimientos a un día.

Propusimos una prueba para la especificación de modelos paramétricos de dos factores. Se presentan tres pruebas diferentes. Las dos primeras se basan en comparar el estimador de la densidad kernel de la función de densidad desconocida y el estimador de $\pi(x, \beta)$ por el método Delta. La última prueba se basa en la idea de comparar el estimador de la densidad kernel y el modelo paramétrico del kernel suavizado para evitar los efectos del sesgo. La ventaja de la última prueba con respecto a las dos primeras es que es válida para datos de bajo-suavizado y datos de sobre-suavizado. La prueba se puede aplicar a muchos procesos de interés en finanzas. En particular, la aplicamos para determinar si la dinámica de la estructura temporal de los CETES se puede modelar a partir de los modelos de Brennan-Schwartz (1979) y Schaefer-Schwartz (1984), los resultados de la prueba nos dicen que ambos modelos continuos son rechazados al nivel del 5% con el fin de describir la dinámica de los CETES. Finalmente analizamos las principales fuentes del rechazo de ambos modelos.

Bibliografía

- [1] Aït-Sahalia, Y. (1996). *Testing Continuous-Time Models of The Spot Interest Rate*. Review of Financial Studies 9: pp. 385-426.
- [2] Aït-Sahalia, Y. (1992). *The Delta Method for Nonparametric Kernel Functionals*. Working paper, MIT.
- [3] Alexander, C. (2003). *Common correlation and calibrating LIBOR*. Wilmott Magazine, March 2003, pp. 68-78.
- [4] Ang, A. and Piazzesi, M. (2003). *A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomics and Latent Variables*. Journal of Monetary Economics, 50, pp. 745-787.
- [5] Bakshi G.S. and Chen Z. (1996). *Inflation, asset prices and the term structure of interest rates in monetary economies*. The Review of Financial Studies 9, pp. 241-275.
- [6] Barber, J. R., Copper, M.L. (1996). *Immunitization Using Principal Component Analysis*. Journal of Portfolio Management, pp. 99-105.
- [7] Bühler, A., Zimmerman, H. (1996). *A Statistical Analysis of the Term Structure of Interest Rates in Switzerland and Germany*. Journal of Fixed Income, 6(3), pp. 55-67.
- [8] Buraschi A, Jiltsov A. (2001). *Time-varying inflation risk premia and the expectations hypothesis: a monetary model of the treasury yield curve*. Journal of Financial Economics in press.

- [9] Chen, L. (1996). *Stochastic mean and stochastic volatility - a three factor model of the term structure of interest rates and its application in derivatives pricing and risk management*. Financial Markets, Institutions and Instruments, 5, pp. 1-87.
- [10] Chen, R.R. and L. Scott. (1993). *Maximum Likelihood Estimation for a Multifactor Equilibrium Model of the Term Structure of Interest Rates*. Journal of Fixed Income, 3, pp. 14-31.
- [11] Cortázar, G. and E. Schwartz (1994). *The valuation of commodity-contingent claims*. Journal of Derivatives, vol. 1, no. 4, pp. 27-39.
- [12] Cortés, J., C. Capistrán, M. Ramos-Francia y A. Torres. (2009). *An Empirical Analysis of the Mexican Term Structure of Interest Rates*. Economics Bulletin, Vol. 29, Núm. 3, pp. 2310-2323.
- [13] Cox, J.C., J.E. Ingersoll, and S.A. Ross. (1985). *A Theory of the Term Structure of Interest Rates*. Econometrica, 53, pp. 385-407.
- [14] Cultberson J. (1957). *The term structure of interest rates*. Quarterly Journal of Economics Vol. 71, pp. 485-517.
- [15] Dai, Q., Singleton, K., 2002. *Expectation puzzles, time-varying risk premia, and affine models of the term structure*. Journal of Financial Economics 63, 415-441.
- [16] D'Ecclesia, R.L., Zenios, S.A. (1994). *Risk Factor Analysis and Portfolio Immunization in the Italian Bond Market*. Journal of Fixed Income 4(2), pp. 51-58.
- [17] Duffie, D. and R. Kan. (1996). *A Yield-Factor Model of Interest Rates*. Mathematical Finance, 6, 4, pp. 379-406.
- [18] Dybvig, P.H. (1989). *Bond and Option Pricing Based on The Current Term Structure*. Working paper, Washington University.
- [19] Elton E.J., M Gruber y R. Michaely (1990). *The Structure of Spot Rates and Immunization*. The Journal of Finance, Vol 45, N. 2, pp. 629-642.

- [20] Engle, R. F. (1982). *Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation*. *Econometrica* 50, pp. 987-1008.
- [21] Fisher I., (1896). *Appreciation and interest*. *AEA Publications* 3(11), pp. 331-442.
- [22] Golub, B., Tilman, M. (1997). *Measuring Yield Curve Risk Using Principal Components Analysis, Value and Risk, and Key Rate Durations*. *Journal of Portfolio Management*, pp. 72-84.
- [23] Heath, R. A., Jarrow, R. A., Morton A. (1992). *Bond pricing and the term structure of interest rates: a new methodology for contingent claim valuations*. *Econometrica*, 60, pp. 77-105.
- [24] Hicks J., (1939). *Value and Capital*. Segunda Edición, Londres: Oxford University Press.
- [25] Jiménez, V. (2002). *Testing the Stability of the Components Explaining Changes of the Yield Curve in Mexico. A Principal Component Analysis Approach*. *Trans 27th ICA 2002*.
- [26] Kanony, C., Mokrane, M. (1992). *Reconstitution de la courbe des taux, analyse des facteurs d'évolution et couverture factorielle*. *Cahiers de la Caisse Autonome de Refinancement* 1, june 1992.
- [27] Karki, J., Reyes, C. (1994). *Model Relationship*. *Risk* 7(12), pp.32-35.
- [28] Knez J., Litterman R. and Scheinkman J. (1996). *Exploration into Factors Explaining Money Markets Returns*. *Journal of Finance*, 49-5, pp. 1861-1881.
- [29] Lardic, S., Priaulet, P., Priaulet, S. (2003). *PCA of the yield curve dynamics: questions of methodologies*. *Journal of Bond Trading and Management*, vol. 1, no. 4, pp. 327-349.
- [30] Lekkos, I.(2000). *A Critique of Factor Analysis of Interest Rates*. *Journal of Derivatives*, Fall 2000, pp. 72-83.
- [31] Litterman, R., Scheinkman, J.A. (1991). *Common Factors Affecting Bond Returns*. *Journal of Fixed Income*, 1, pp. 54-61.

- [32] Lund J. (1997). *Non-Linear Kalman Filtering Techniques for Term Structure Models*. Aarhus School of Business.
- [33] Mardia, K.V., Kent, J.T., Bibby, J.M. (1992). *Multivariate Analysis*. New York: Academic Press.
- [34] Martinelli, L., Priaulet, P., Priaulet, S. (2003). *Fixed income securities*. New York: Wiley, 1997.
- [35] Modigliani, F., Sutch, R. (1966). *Innovations in interest rate policy*. American Economic Review, Papers and Proceedings Supplement. Vol. 56, pp. 178-97.
- [36] Penachi, G.G. (1991). *Identifying the Dynamics of Real Interest Rates and Inflation: Evidence using Survey Data*. Review of Financial Studies, 4, pp. 53-86.
- [37] Pearson, N.D., Sun, T.S. (1994). *Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross Model*. Journal of Finance, 49, pp. 1279-1304.
- [38] Singh, M.K. (1997). *Value-at-Risk Using Principal Components Analysis*. The Journal of Portfolio Management, 24, pp. 101-112.
- [39] Steeley, J.M. (1990). *Modelling the dynamics of the term structure of interest rates*. Economic and Social Review, vol. 21, no. 4, pp. 337-361.
- [40] Steeley, J. (1991). *Estimating The Gilt-Edged Term Structure: Basis Splines and Confidence Internals*. Journal of Business Finance and Accounting 18: pp. 513-529.
- [41] Steeley, J. (1991). *Estimating The Gilt-Edged Term Structure: Basis Splines and Confidence Internals*. Journal of Business Finance and Accounting 18: pp. 513-529.
- [42] Svensson, L. (1994). *Estimating and interpreting forward interest rates: Sweden 1992-94*. CEPR Discussion Paper 1051.
- [43] Vasicek, O.A. (1997). *An equilibrium Characterization of the Term Structure*. Journal of Financial Economics, 5, pp. 177-188.

- [44] Villarreal, A. (2001). *An Empirical Analysis of the Mexican Term Structure of Interest Rates and some evidence for other Latin American countries*. Disertación doctoral, Tulane University.