



Universidad
Nacional
de Córdoba



Facultad de Matemática,
Astronomía, Física y
Computación

TESIS DOCTORAL

Modelo de cointegración multifactorial aplicado a la curva de swap spread

AUTOR: ELISA RAVASI¹

DIRECTORA: DRA. NOEMÍ P. KISBYE²

FACULTAD DE MATEMÁTICA, ASTRONOMÍA, FÍSICA Y COMPUTACIÓN,
UNC

Comisión Asesora:

Dr. Oscar H. Bustos³ (Codirector)

Dr. Fernando Menzaque⁴



Esta obra está bajo una Licencia Creative Commons Atribución - No Comercial - Sin Obra Derivada 4.0 Internacional.

^{1,2,3,4}Facultad de Matemática, Astronomía, Física y Computación-UNC, Argentina.

¹elisa.ravasi@unc.edu.ar

²patricia.kisbye@unc.edu.ar

Resumen

En esta tesis se trabajó en cómo capturar los movimientos estructurales de la curva Swap Spread a través de un modelo dinámico de factores en el que se asume que cada factor: nivel, pendiente y curvatura están cointegrados con variables macroeconómicas que los impulsan: los valores respaldados por hipotecas (MBS por sus siglas en inglés) y el tipo de cambio dólar/yen (USD/JPY).

Se llevó a cabo una comparación entre un modelo autorregresivo vectorial (VAR) compuesto por factores y drivers, y un modelo de corrección de errores vectorial (VECM) que asume que esos factores y drivers son integrados y hay una o más relaciones de cointegración entre ellos. La comparación se realizó simulando las series de tiempo para cada modelo y luego estudiando su eficiencia en la predicción de cada factor. Mostramos que asumir la integrabilidad de los factores y drivers así como una cointegración entre estas variables mejora el desempeño del modelo original (VAR).

Finalmente se vio que este enfoque también permite estudiar el impacto que diferentes drivers pueden tener sobre un factor en particular más que en cada punto de la curva, dando una comprensión más global del comportamiento de la curva Swap Spread.

Palabras clave: Curva swap spread, Macroeconomía, Nelson y Siegel dinámico, cointegración.

Código:

Agradecimientos

Quiero comenzar agradeciendo a la FaMAF por haberme dado la posibilidad de desarrollar mi doctorado en esta área y por la excelente formación que me permitió abordar esta disciplina de la mejor manera.

Quiero dar las gracias al personal de posgrado especialmente a Celeste y a Daniel por su gran trabajo y disposición siempre.

También quiero destacar y agradecer las becas otorgadas inicialmente por Secyt y Conicet.

Agradezco a la empresa Arfima Trading por orientarme al comienzo de esta tesis en el planteo del problema y la bibliografía inicial. Además, por brindarme los datos necesarios para desarrollar la aplicación.

Estoy infinitamente agradecida a todos los docentes de FaMAF no solo por la formación académica brindada sino también por la paciencia y dedicación. Especialmente a Oscar Bustos por todo el conocimiento transmitido y por el gran esfuerzo en entender el trabajo y ayudarme a mejorarlo.

Agradezco a mi comisión asesora por apoyarme durante todo el doctorado así como al tribunal de esta tesis por la dedicación y el trabajo realizado en la corrección de la misma.

Quiero agradecer a Marco y Fede de Quanticko por el inmenso apoyo sin el cual hubiera sido imposible terminar este trabajo.

A mi mamá, a mi papá y a mis hermanas por el amor inmenso e incondicional. Por estar ahí y ayudarme siempre. A Mónica por su gran apoyo y cariño.

A los hermosos amigos que me dejó el doctorado: Kar, gracias por todo el cariño, el aguante. Por transmitirme toda tu energía y empuje. A Vane, mi gran amiga, gracias por estar siempre. A Elena, Romi, Kari, Dami gracias por los almuerzos, las risas y los consejos. A mis amigos de la vida por su apoyo y amor infinito. Mi amiga y hermana Romi. A Gaby, Fer, Romi y toda la banda gracias por estar siempre!

Quiero agradecer a Pablo, mi compañero en la vida por hacer posible lo imposible. Por su amor, su apoyo y su paciencia. Porque sin él nada de esto hubiera sido posible. A Brei, mi hijo hermoso y maravilloso por darle sentido a todo. Gracias! Los amo!

Por último, quiero agradecer a la persona más importante en este proceso: a Patri. No tengo palabras para agradecer lo que me ha dado siempre. Desde el primer día que toqué la puerta de su oficina para preguntarle si se embarcaba conmigo en esto hasta el último mail confirmando el día de defensa. Gracias!! Gracias por no bajar los brazos nunca. Gracias por enseñarme y quererme tanto!!

Índice general

Abstract	3
Agradecimientos	5
Nomenclatura	11
Lista de Tablas	12
Lista de Figuras	15
Introducción	17
1. Conceptos financieros sobre la estructura temporal de tasas de interés	21
1.1. Tasas de interés	21
1.1.1. Yields y bonos con cupón	25
1.2. Estructura temporal de tasas de interés	28
1.3. Swap spread	30
1.3.1. Contratos futuros IRS: Interest Rate Swap	30
1.3.2. Curva de swap spread	31
2. Modelos de Nelson y Siegel para la estructura temporal de tasas de interés	35
2.1. Introducción	35
2.2. Descripción del modelo de Nelson y Siegel estático	36
2.3. Descripción del modelo de Nelson y Siegel dinámico DNS	37
2.3.1. Modelos Macro-Financieros	41
3. Cointegración y modelo de corrección de error	45
3.1. Procesos autorregresivos y condición de estabilidad	45
3.1.1. Procesos VAR(1)	46
3.1.2. Procesos VAR(p) y la condición de estabilidad.	49

3.1.3.	Procesos estacionarios	51
3.1.4.	Procesos estacionarios con tendencia	53
3.2.	Procesos integrados	54
3.3.	Procesos cointegrados y modelo de corrección de error	59
3.3.1.	Modelo de corrección de error	60
3.3.2.	Modelo de corrección de error con término determinista	64
3.3.3.	Teorema de Representación de Granger	66
3.3.4.	Análisis del término determinista en la Representación de Granger	77
3.4.	Estimación del modelo de corrección de error	81
3.4.1.	Estimación ML	82
4.	Modelo de cointegración multifactorial	85
4.1.	Introducción	85
4.2.	Especificación del modelo VAR aplicado a DNS y drivers (VAR DNS)	88
4.3.	Especificación del modelo VECM aplicado a DNS y drivers (VECM DNS)	89
4.4.	Comparación entre los modelos VAR y VECM aplicados a DNS y drivers	92
4.5.	Estimación de los modelos	95
4.5.1.	Datos y estimación del modelo VAR DNS	95
4.5.2.	Análisis inicial y estimación del modelo VECM DNS	96
4.5.3.	Implementación en Python	104
4.6.	Simulación y resultados	106
4.6.1.	Caso A: Nivel y MBS	106
4.6.2.	Caso B: Pendiente y USD/JPY	112
4.6.3.	Caso C: Curvatura y MBS	117
4.6.4.	Resumen de resultados	124
5.	Conclusiones y futuras líneas de investigación	127
5.1.	Conclusiones	127
5.2.	Futuras líneas de investigación	128
A.	Modelos de tasa de interés	131
A.1.	Conceptos generales del modelado de tasas de interés	131
A.1.1.	Proceso de precios de bonos y modelado general de tasa short	131
A.1.2.	Expectativas, prima de riesgo y convexidad	133
A.2.	Modelos de un factor: tasa short	137
A.2.1.	Modelos de equilibrio	137
A.2.2.	Modelos libre de arbitraje	139

<i>ÍNDICE GENERAL</i>	9
A.3. Modelos multifactoriales	141
A.3.1. Modelos afines	142
A.3.2. Modelos de tasa forward	144
B. Análisis sobre los parámetros λ y Λ en el modelo de Nelson y Siegel	147
B.1. Análisis del parámetro λ	147
B.2. Propiedad de la matriz Λ	153
C. Descripción de test estadísticos	159
C.1. Test lag óptimo: Criterios de información	159
C.2. Test raíz unitaria: Dickey-Fuller Aumentado	160
C.3. Cointegración: Test de Johansen	161
C.3.1. Términos deterministas	162
D. Resultados: predicción de 2 pasos	165
D.1. Caso A: Nivel y MBS	165
D.1.1. Modelo VAR	165
D.1.2. Modelo VECM	167
D.2. Caso B: Pendiente y USD/JPY	168
D.2.1. Modelo VAR	168
D.2.2. Modelo VECM	169
D.3. Caso C: Curvatura y MBS	170
D.3.1. Modelo VAR	170
D.3.2. Modelo VECM	171
Bibliografía	173

Nomenclatura

ADF	Test de Dickey-Fuller aumentado.
AFNS	Modelo de Nelson y Siegel libre de arbitraje.
AIC	Akaike information criterion.
AR	Modelo autorregresivo unidimensional.
BIC	Bayesian Information Criteria.
CIR	Modelo de Cox, Ingersoll y Ross
CME	Chicago Mercantile Exchange.
CMT	Constant Maturity Treasury.
DF	Test de Dickey-Fuller.
DNS	Modelo de Nelson y Siegel dinámico.
ECM	Modelo de Corrección de Error.
FRA	Contrato forward-rate.
GDP	Gross Domestic Product.
HJM	Modelo de Heath, Jarrow y Morton
HQ	Hannan–Quinn information criterion.
IRS	Interest Rate Swap.
LIBOR	London InterBank Offered Rate.
Libor-Repo O/N	Diferencia entre la tasa Libor y la tasa Repo Overnight.

LR	Likelihood Ratio
MBS	Effective Duration de los valores respaldados por hipotecas (Mortgage backed securities).
ML	Máxima verosimilitud.
MSE	Error Cuadrático Medio.
OLS	Ordinary least squares
S&P	Standard and Poor's
SC	Schwarz information criterion.
Slope Tsy 2/10	Diferencia entre la tasa del Tesoro de EEUU con vencimiento a 10 años y la tasa a 2 años.
TED spread	Diferencia entre la tasa interbancaria a 3 meses en dólares y la tasa del Tesoro de EEUU a 3 meses.
USD / JPY	Tipo de cambio entre el Yen y el Dólar.
VAR	Modelo autorregresivo vectorial.
VECM	Modelo de corrección de errores vectorial.

Índice de tablas

4.1. Estadísticos de factores y drivers	98
4.2. Test ADF de factores y drivers- Sin asumir tendencia temporal en la media.	100
4.3. Casos considerados para el test de Johansen, Johansen [35]	106
4.4. Resultados del modelo VAR que incluye el factor 1 o nivel y driver MBS con datos diarios desde el 25/08/2011 hasta el 04/08/2014.	108
4.5. Resultados del modelo VECM que incluye el factor 1 o nivel y el driver MBS con datos diarios desde el 25/08/2011 hasta el 04/08/2014 asumiendo intercepto y tendencia fuera de la ecuación de cointegración.	110
4.6. Test ADF (ct) para VECM - Caso A	110
4.7. Test ADF (nc) - Caso A	110
4.8. Resultados del VAR que incluyen el factor 2 o pendiente y el driver USD/JPY con data diaria desde 28/03/2013 to 13/08/2015.	114
4.9. Resultados de VECM que incluyen el factor 2 o pendiente y el driver USD/ JPY con datos diarios desde el 28/03/2013 hasta el 13/08/2015 asumiendo intercepto y tendencia fuera de la ecuación de cointegración.	116
4.10. Test ADF (ct) para VECM - Caso B	116
4.11. Test ADF (nc) - Caso B	118
4.12. Resultados del VAR que incluye el factor 3 o curvatura y el driver MBS con datos diarios desde 12/06/2012 a 15/08/2014.	119
4.13. Resultados de VECM que incluye el factor 3 o curvatura y el driver MBS con datos diarios desde 12/06/2012 al 15/08/2014.	121
4.14. Test ADF (ct) para VECM - Caso C	123
4.15. Test ADF (nc) - Caso C	124
4.16. Caso A: Resumen resultados.	124
4.17. Caso B: Resumen resultados.	124
4.18. Caso C: Resumen resultados.	125

D.1. Resultados del modelo VAR que incluye el factor 1 o nivel y driver MBS con datos diarios desde el 25/08/2011 hasta el 04/08/2014.	166
D.2. Resultados del modelo VECM que incluye el factor 1 o nivel y el driver MBS con datos diarios desde el 25/08/2011 hasta el 04/08/2014 asumiendo intercepto y tendencia fuera de la ecuación de cointegración.	167
D.3. Resultados del VAR que incluye el factor 2 o pendiente y el driver USD/JPY con data diaria desde 28/03/2013 to 13/08/2015.	168
D.4. Resultados de VECM que incluye el factor 2 o pendiente y el driver USD/ JPY con datos diarios desde el 28/03/2013 hasta el 13/08/2015 asumiendo intercepto y tendencia fuera de la ecuación de cointegración.	169
D.5. Resultados del <i>VAR</i> que incluye el factor 3 o curvatura y el driver MBS con datos diarios desde 12/06/2012 a 15/08/2014.	170
D.6. Resultados del modelo VECM que incluye el factor 3 o curvatura y el driver MBS con datos diarios desde 12/06/2012 al 15/08/2014 asumiendo intercepto y tendencia fuera de la ecuación de cointegración.	171

Índice de figuras

1.1. Curva de rendimiento del Tesoro Americano. Fuente: U.S. Department Of The Treasury (https://www.treasury.gov)	29
1.2. Curva o Estructura temporal de tasas Swap Spread. Fuente: Bloomberg. Datos facilitados por la empresa Arfima.	33
1.3. Serie de tiempo de tasas Swap Spread. Fuente: Bloomberg. Datos facilitados por la empresa Arfima.	34
3.1. Series de tiempo bivariadas generadas por procesos estacionarios.	52
3.2. Series de tiempo bivariadas generadas por procesos no estacionarios.	52
3.3. Ejemplos de caminata aleatoria.	55
3.4. Ejemplo de caminata aleatoria con drift.	56
3.5. Ejemplo de Procesos Cointegrados.	59
4.1. Las curvas de color azul corresponden a las observadas y las de color rojo a las estimadas con el método de Nelson y Siegel	96
4.2. Serie del factor nivel y serie de sus diferencias	99
4.3. Serie del factor pendiente y serie de sus diferencias	99
4.4. Serie del factor curvatura y serie de sus diferencias	99
4.5. Series del nivel y drivers:	101
4.6. Series de la pendiente y drivers:	102
4.7. Series de la curvatura y drivers:	103
4.8. <i>Factor 1 y driver MBS</i> - Período: 25/08/2011 a 04/08/2014.	107
4.9. <i>Trayectoria del Factor 1-Caso A (VAR):</i> Muestra Real (Factor Nivel) y última simulación (Nivel VAR).	109
4.10. <i>Trayectoria del Factor 1-Caso A (VECM):</i> Muestra Real (Factor Nivel) y última simulación (Nivel VECM).	111
4.11. <i>Trayectoria $\beta\tilde{X}_t$:</i> Tendencia (Factor Nivel-MBS) y su diferencia.	112
4.12. <i>Factor 2 and driver USD/JPY</i> - Período: 28/03/2013 to 13/08/2015.	113

4.13. <i>Trayectoria del Factor 2-Caso B (VAR): Muestra Real (Factor Pendiente) y última simulación (VAR Pendiente).</i>	115
4.14. <i>Trayectoria del Factor 2-Case B (VECM): Muestra Real (Factor Pendiente) y última simulación (VECM Pendiente).</i>	117
4.15. <i>Trayectoria de $\beta\tilde{X}_t$: Tendencia (Factor Pendiente-JPY) y su serie de diferencias.</i>	118
4.16. <i>Factor 3 y driver MBS - Período: 12/06/2012 al 15/08/2014.</i>	119
4.17. <i>Trayectoria del Factor 3-Case C (VAR): Muestra Real (Factor Curvatura) y última simulación (VAR Curvatura).</i>	120
4.18. <i>Trayectoria del Factor 3-Caso C (VECM): Muestra Real (Factor Curvatura) y última simulación (VECM Curvatura).</i>	122
4.19. <i>Trayectoria de $\beta\tilde{X}_t$: Tendencia (Factor Curvatura-MBS) y su serie de diferencias.</i> . . .	123
B.1. <i>Gráfico de las cargas para dos valores distintos de λ.</i>	148
B.2. <i>Ajuste con Nelson y Siegel</i>	149

Introducción

En la actualidad, hay cada vez más activos financieros relacionados a las variaciones de curvas de tasas de interés. Comprender la evolución dinámica de dichas curvas es importante, entre otras cosas, para poder valorar precios de activos financieros y sus derivados, el manejo de riesgos financieros, la estructuración de la deuda fiscal y la conducción de las políticas monetarias. Por ello, se han realizado numerosas investigaciones y se ha producido una vasta literatura a los fines de obtener un mayor entendimiento de la dinámica de tales curvas de rendimiento.

Una curva de tasas que ha sido ampliamente estudiada y analizada en los últimos años es la correspondiente a los Swaps Spread. Esta curva, que detallaremos a lo largo de la tesis, está ampliamente ligada al riesgo de crédito y es un posible indicador de crisis financieras. De ahí su gran importancia.

Si bien en el presente trabajo se abordará un problema multidisciplinario, el lector puede comprender los detalles del objeto de estudio en lo desarrollado a través del primer Capítulo: “Conceptos financieros sobre la estructura temporal de tasas de interés”. Sin embargo, es esperable que se manejen conceptos básicos de probabilidad y estadística como series de tiempo, procesos autorregresivos univariados y test estadísticos.

Las ideas detrás de la valuación de derivados sobre acciones y renta fija son similares pero existen importantes diferencias. En muchas aplicaciones de valuación de activos, las tasas de interés se asumen como una función determinística del tiempo. Esto está motivado usualmente por el hecho que la variabilidad de dichas tasas contribuyen en mucho menor orden al valor del derivado que la fluctuación de los activos subyacentes. Sin embargo, cuando consideramos un derivado cuyo valor al vencimiento (payoff) depende de alguna manera del nivel de las tasas de interés, tiene sentido descartar la hipótesis determinística y construir un modelo en el cual dichas tasas se basen en un proceso aleatorio o estocástico.

A través de los años, los investigadores han desarrollado una amplia variedad de modelos matemáticos que buscan explicar y comprender diversas curvas de rendimientos. Algunos de ellos son teóricamente rigurosos pero deficientes a nivel empírico. Otros manifiestan un buen desempeño empírico pero suelen dejar de lado ciertos aspectos teóricos en los que se basa la teoría financiera. Por último, varios de esos modelos no tienen en cuenta los diversos factores macroeconómicos que gobiernan tales curvas.

En este trabajo, nuestro objetivo es construir un modelo matemático que permita comprender los mo-

vimientos de la curva de Swap Spread de manera estructural, que simultáneamente manifieste tratabilidad empírica, buen ajuste e incorpore variables macroeconómicas directamente al modelo.

Si bien en los últimos años se han desarrollado modelos que incorporan variables fundamentales, como es el caso del trabajo de Kobor et al. [36], el modelo obtenido arroja resultados relacionados con 3 puntos particulares de la curva de Swap Spread correspondiente a los vencimientos a 2, 5 y 10 años. Sin embargo, un desafío de gran interés es intentar captar los movimientos de una manera más global a lo largo de la curva. Además, buscamos brindarle a nuestro modelo un marco de mayor formalidad matemática. Por otro lado, es importante que continúe siendo parsimonioso, es decir, que capte la compleja naturaleza de los datos a través del menor número posible de parámetros, que sea simple y que incorpore los drivers directamente al modelo. Esto se aleja de trabajos como el de Feldhütter and Lando [27] donde si bien se presenta una estructura formal e imposiciones sobre libre arbitraje, también se plantea un modelo de tipo afín complejo de 6 factores y donde no se incluyen los drivers directamente al modelo.

Un modelo parsimonioso, que produce buenas predicciones y que es flexible para adoptar diversas formas de la curva de rendimientos es el introducido por Nelson and Siegel [47]. Estos autores describen la curva de tasas forward instantáneas con una forma funcional de la cual se pueden extraer 3 variables: L , S y C que pueden ser interpretadas como tres factores latentes. Estos factores se suelen denominar “Nivel”, “Pendiente” y “Curvatura” ya que se asocian a relaciones o características estructurales de las curvas de tasas de interés (ver Diebold and Li [18]). A su vez, artículos como el de Diebold et al. [20], permiten a estos factores variar en el tiempo y estudian las series temporales $\{L_t\}$, $\{S_t\}$ y $\{C_t\}$ como procesos autorregresivos. Los autores del artículo previamente mencionado, también buscan caracterizar las relaciones entre $\{L_t\}$, $\{S_t\}$ y $\{C_t\}$ y la macroeconomía. Como medidas de la economía incluyen tres variables clave: la utilización de la capacidad de fabricación, la tasa de fondos federales y la inflación anual de los precios. Luego, estiman una extensión del modelo de Nelson y Siegel para la curva de tasas del Tesoro de EE.UU. considerando 6 series de tiempo asociadas a nivel, pendiente y curvatura junto con las tres variables macroeconómicas. Es decir, obtienen $\{L_t\}$, $\{S_t\}$, $\{C_t\}$, $\{CU_t\}$, $\{FFR_t\}$ y $\{INFL_t\}$ a partir de Nelson y Siegel dinámico. Sin embargo, asumen un modelo estacionario tanto para los factores como para las variables económicas a pesar de que se concluye lo contrario a través de diversos tests estadísticos como el test de Dickey-Fuller (DF) o Dickey-Fuller aumentado (ADF) entre otros.

Por lo tanto, para llevar a cabo nuestro objetivo y capturar las dinámicas de la curva utilizaremos un modelo de estructura a término de tres factores dinámico asumiendo integrabilidad para cada uno y cointegración con los drivers que los conducen. Es decir, primero analizaremos un modelo vectorial autorregresivo (VAR) aplicado a la curva de Swap Spread junto con los drivers que la conducen. Luego, descartaremos la hipótesis de estacionariedad para incorporar las relaciones de cointegración entre factores y drivers asumiendo un modelo vectorial de corrección de error (VECM).

Este modelo tiene la flexibilidad necesaria para adaptarse a la forma cambiante de la curva de rendimiento además de ser parsimonioso y fácil de estimar. Mostraremos que la incorporación de integrabilidad y la suposición de cointegración entre las variables no sólo mantiene la parsimonia sino que mejora el desempeño del modelo original. Además, permite determinar relaciones lineales explícitas a corto y largo plazo entre los factores que definen la curva y los drivers económicos.

Analizaremos cointegración entre los factores latentes de nivel, pendiente y curvatura con las variables macroeconómicas que se plantean en el artículo de Kobor et al. [36] del Banco Mundial: “What Determines U.S. Swap Spreads?”. donde se detallan los drivers de mercado que afectan a dicha curva. A esto lo llevaremos a cabo utilizando la teoría desarrollada por Johansen y que describimos en el Capítulo 3. Esto no sólo le dará profundidad matemática y evitará considerar un modelo estacionario sino que servirá para encontrar relaciones a corto y largo plazo entre los factores que definen la curva y los drivers económicos.

En el Capítulo 1 desarrollaremos la teoría financiera necesaria para comprender las características fundamentales de nuestro objeto de estudio: curvas de tasas de interés. Además, esto nos servirá para abordar el Capítulo 2 en el que expondremos el modelo de tasas de Nelson y Siegel y los modelos macroeconómicos existentes. En el Capítulo 3 nos focalizaremos en detallar la teoría matemática que sustentará nuestro modelo. Allí abordaremos en profundidad los conceptos de estabilidad, integración y cointegración así como los Teoremas de Representación de Granger y el modelo de corrección de error vectorial. Para ello, seguiremos principalmente el libro de Lütkepohl [43]. Por último, en el Capítulo 4, plantearemos nuestro modelo y llevaremos a cabo un estudio de simulación que analiza la eficiencia predictiva en relación a un modelo VAR que no incorpora cointegración como el planteado en Diebold et al. [20].

Capítulo 1

Conceptos financieros sobre la estructura temporal de tasas de interés

Comenzaremos detallando las definiciones y conceptos necesarios para construir el marco teórico financiero de las curvas de tasas de interés. Las primeras nociones de tasas de interés, curva de rendimiento y otros estarán guiadas por la estructura provista en Brigo y Mercurio [7]. Una vez introducidos estos conceptos, daremos la definición de curva de tasas swap y swap spread.

1.1. Tasas de interés

Llamamos *cuenta bancaria o cuenta de mercado de moneda* a una inversión libre de riesgo donde el beneficio se acumula continuamente a la tasa libre de riesgo que prevalece en el mercado en cada instante.

Denotamos $B(t)$ al valor de una cuenta bancaria en tiempo $t \geq 0$. Supondremos que la cuenta bancaria evoluciona de acuerdo a la siguiente ecuación diferencial:

$$dB(t) = r(t)B(t)dt, \quad B(0) = 1,$$

donde r es una función positiva del tiempo. Como una consecuencia,

$$B(t) = \exp\left(\int_0^t r(s)ds\right).$$

En la definición que acabamos de brindar, $r(s)$ es la *tasa instantánea* a la cual se acumula la cuenta bancaria. Esta tasa instantánea usualmente se denomina *tasa spot instantánea*, o brevemente, *tasa short*.

Definición 1.1.1. Factor de descuento. Sean t y T dos instantes de tiempo con $t \leq T$. El factor de descuento $D(t, T)$ es la cantidad de moneda en tiempo t que es “equivalente” a una unidad de moneda

que se paga en tiempo T , y es dada por

$$D(t, T) = \frac{B(t)}{B(T)} = \exp\left(-\int_t^T r(s)ds\right).$$

En valoración de derivados, r se asume como una función determinista del tiempo y la variabilidad principal considerada en el modelo está dada por el comportamiento del subyacente: acciones, tipos de cambio, entre otros. Luego, tanto la cuenta bancaria como los factores de descuento resultan ser funciones deterministas del tiempo en cualquier tiempo futuro. Sin embargo, cuando consideramos productos que dependen de la tasa de interés, la variabilidad principal que interesa es claramente la de las tasas de interés en sí mismas. Por lo tanto, en estos casos es necesario dejar de lado la hipótesis determinística y comenzar modelando la evolución de dichas tasas en el tiempo a través de un proceso estocástico $\{r_t\}_{t \geq 0}$. Como una consecuencia, la cuenta bancaria y los factores de descuento serán también procesos estocásticos.

Definición 1.1.2. Bono cupón cero. *Un bono cupón cero de madurez o vencimiento T es un contrato que garantiza a su poseedor el pago de una unidad de moneda en tiempo T , sin pagos intermedios. El valor del contrato en tiempo $t \leq T$ se denota por $P(t, T)$. Se cumple que $P(t, T) > 0$ y $P(T, T) = 1$.*

La diferencia entre el factor de descuento $D(t, T)$ y el precio del bono cupón cero $P(t, T)$ cae en que los dos objetos son respectivamente una “cantidad de moneda equivalente” y un “valor de un contrato”. Sin embargo, ambos valores deben coincidir, de lo contrario habría posibilidad de una ganancia libre de riesgo. Esto último se conoce como una posibilidad de *arbitraje*.

Los precios de los bonos cupón cero son las cantidades básicas en la teoría de tasas de interés, y todas las tasas de interés se pueden definir en términos de precios de bonos cupón cero como veremos ahora. Notemos, sin embargo, que las tasas de interés son las que usualmente cotizan en mercados financieros (interbancarios), mientras que los bonos cupón cero son instrumentos teóricos que, como tales, no son directamente observables en el mercado.

En el pasaje de bonos cupón cero a tasas y viceversa necesitamos conocer dos rasgos fundamentales de las tasas en sí mismas: la clase o tipo de capitalización y la convención day-count que se aplicarán en la definición de tasa. El tipo de capitalización viene dado por la manera en que se va a acumular una inversión vigente en tiempo t con madurez T a través de la tasa de interés. La convención day-count se refiere a la forma de medir la fracción de año entre t y T . A esta distancia temporal medida en años la denotaremos por $\tau(t, T)$. Por ejemplo, la mayoría de los depósitos monetarios y notas con tasa variable se calcula en una convención day-count real/360, mientras que los bonos y notas emitidos por el Tesoro de EE.UU. generan intereses calculados sobre una base real/real. Para más detalle sobre estos conceptos, ver Brigo and Mercurio [7].

Así, la *tasa de interés spot capitalizada continuamente* $R(t, T)$ en tiempo t para la madurez o vencimiento T es la tasa constante a la cual una inversión de $P(t, T)$ unidades de moneda en tiempo t acumula continuamente produciendo una cantidad unitaria de moneda en la madurez T .

En fórmulas:

$$R(t, T) \doteq -\frac{\ln P(t, T)}{\tau(t, T)}.$$

Notemos que la tasa de interés capitalizada continuamente es por lo tanto una tasa constante que es consistente con los precios de los bonos cupón cero en el sentido que:

$$e^{R(t, T)\tau(t, T)} P(t, T) = 1.$$

De aquí que podemos expresar el precio del bono en términos de la tasa capitalizada continuamente R :

$$P(t, T) = e^{-R(t, T)\tau(t, T)}.$$

La convención day-count involucrada en la capitalización continua es usualmente $\tau(t, T) = T - t$, la diferencia de tiempo expresada en años.

De manera similar, la *tasa de interés capitalizada simplemente en tiempo t para la madurez o vencimiento T* , que denotaremos por $L(t, T)$, es la tasa constante que produce una acumulación o rendimiento proporcional al tiempo de inversión. Es decir,

$$L(t, T) \doteq \frac{1 - P(t, T)}{\tau(t, T)P(t, T)} = \frac{1}{\tau(t, T)} \left(\frac{1}{P(t, T)} - 1 \right). \quad (1.1)$$

Por ejemplo, las tasas LIBOR (London InterBank Offered Rate) de mercado son tasas de interés ampliamente utilizadas y se capitalizan simplemente (lo cual motiva el hecho de denotar por L a tales tasas).

La definición recién dada de $L(t, T)$ conduce inmediatamente al siguiente resultado:

$$P(t, T)(1 + L(t, T)\tau(t, T)) = 1,$$

de modo que el precio de un bono puede expresarse en términos de L como:

$$P(t, T) = \frac{1}{1 + L(t, T)\tau(t, T)}.$$

Cabe señalar que existen otros tipos de capitalizaciones como la compuesta anual, k veces por año, entre otras.

Finalmente, daremos la definición de tasas forward las cuales están caracterizadas por tres instantes de tiempo t , T y S tal que $t < T < S$. El tiempo t es en el cual se considera la tasa correspondiente para el período $[T, S]$. El tiempo T se denomina expiración y el tiempo S es la madurez. Las tasas forward son tasas de interés que se fijan en un instante t para una inversión en un período de tiempo futuro, y que se ajustan consistentemente con la estructura a término de los factores de descuento.

Definición 1.1.3. *Tasa de interés forward capitalizada simplemente.* La tasa de interés forward capitalizada simplemente vigente en tiempo t para la expiración $T > t$ y la madurez o vencimiento $S > T$ se denota por $F(t; T, S)$ y está definida por

$$F(t; T, S) \doteq \frac{1}{\tau(T, S)} \left(\frac{P(t, T)}{P(t, S)} - 1 \right).$$

Si consideramos un contrato FRA (forward-rate agreement) que da a su poseedor un pago de intereses para el período comprendido entre T y S , dicha tasa forward se puede ver como la tasa fija K en tal contrato con expiración T y madurez S que hace al FRA un contrato justo en tiempo t . Para más detalle ver Brigo and Mercurio [7]

La tasa forward $F(t; T, S)$ puede considerarse como un estimador de la tasa spot futura $L(T, S)$, la cual es incierta en tiempo t , basada en las condiciones de mercado en tiempo t (ver fórmula (1.1)). Cuando la madurez de la tasa forward tiende a su expiración, nos referimos a la noción de *tasa forward instantánea*.

Definición 1.1.4. *Tasa de interés forward instantánea.* La tasa de interés forward instantánea vigente en el tiempo t para la madurez $T > t$ se denota por $f(t, T)$ y se define como

$$f(t, T) \doteq \lim_{S \rightarrow T^+} F(t; T, S). \quad (1.2)$$

Luego, si suponemos que la función $T \mapsto P(t, T)$ es derivable y calculamos el límite (1.2), obtenemos:

$$\begin{aligned} \lim_{S \rightarrow T^+} F(t; T, S) &= - \lim_{S \rightarrow T^+} \frac{1}{P(t, S)} \frac{P(t, S) - P(t, T)}{S - T} \\ &= - \frac{1}{P(t, T)} \frac{\partial P(t, T)}{\partial T} \\ &= - \frac{\partial \ln P(t, T)}{\partial T}, \end{aligned}$$

ya que cerca del límite utilizamos la convención $\tau(T, S) = T - S$. Entonces,

$$f(t, T) = - \frac{\partial \ln P(t, T)}{\partial T},$$

de modo que

$$P(t, T) = \exp \left(- \int_t^T f(t, u) du \right). \quad (1.3)$$

Intuitivamente, la tasa de interés forward instantánea $f(t, T)$ es una tasa de interés forward en tiempo t cuya madurez o vencimiento está muy cerca de su expiración T , digamos $f(t, T) \approx F(t; T, T + \Delta T)$ con ΔT pequeño. Las tasas forward instantáneas son cantidades fundamentales en la teoría de tasas de interés. De hecho, resulta que una de las formas más generales de imponer “ausencia de oportunidades

de arbitraje” en un modelo de tasa de interés está relacionando ciertas cantidades en la expresión de la evolución de f . El marco teórico que se desarrolla en torno a esta tasa es el de Heath, Jarrow and Morton el cual describimos en A.3.2 basándonos en Heath et al. [31].

En general, es razonable definir la tasa spot instantánea como la tasa forward instantánea cuando la madurez T tiende a t , es decir:

Definición 1.1.5. *Sea t un instante de tiempo tal que $0 \leq t \leq T$, $f(t, T)$ la tasa de interés forward instantánea vigente en el tiempo t para la madurez o vencimiento T y sea $r(t)$ la tasa spot instantánea en tiempo t . Entonces:*

$$r(t) \doteq \lim_{T \rightarrow t^+} f(t, T) \doteq f(t, t).$$

Luego, si queremos construir un modelo para el mercado de bonos, podemos hacerlo de muchas maneras diferentes:

- Podemos especificar la dinámica de la tasa short (y luego, quizá tratar de derivar los precios de los bonos utilizando argumentos de arbitraje).
- Podemos especificar directamente la dinámica de todos los posibles bonos.
- Podemos especificar la dinámica de todas las posibles tasas forward, y luego usar (1.3) para obtener precios de bonos.

En el siguiente Capítulo detallaremos estas posibilidades en relación al modelado del mercado de bonos y tasas de interés.

1.1.1. Yields y bonos con cupón

En la mayoría de los mercados de bonos, sólo hay un número relativamente pequeño de bonos cupón cero que se negocian activamente. Los vencimientos suelen ser generalmente cortos (entre medio año y dos años), mientras que la mayoría de los bonos con un tiempo más largo hasta el vencimiento son bonos que pagan cupones. A pesar de este hecho empírico, seguiremos asumiendo la existencia de un mercado para todos los posibles bonos cupón cero, y a continuación introduciremos y valoraremos los bonos con cupón en términos de bonos cupón cero.

Definición 1.1.6. Bono con cupón. *Un bono con cupón es un contrato que paga a su poseedor las cantidades c_1, \dots, c_m en las fechas $T_1 < \dots < T_m$. Luego, el precio del bono que denotamos por $P_C(t)$ en el tiempo t tal que $T_0 \leq t \leq T_k$ y donde T_0 , la fecha de emisión del bono, puede expresarse como la*

suma de los cash flows c_k, \dots, c_m descontados en tiempo t , es decir,

$$P_C(t) = \begin{cases} \sum_{j=k}^m c_j D(t, T_j), & \text{si } T_{k-1} < t \leq T_k, \quad \forall k \geq 2; \\ \sum_{j=1}^m c_j D(t, T_j), & \text{si } T_0 \leq t \leq T_1. \end{cases}$$

Es claro que poseer un bono con cupón es equivalente a poseer un portfolio de bonos cupón cero cada uno con vencimientos o madurez en T_1, \dots, T_m . Más precisamente, es equivalente a tener una cantidad c_k de bonos cupón cero con vencimiento T_k para cada $k = 1, \dots, m$.

Un bono real normalmente paga un cupón c fijo y repaga el principal N en T_m . En tal caso, tenemos $c_j = c$ para $j = 1, \dots, m - 1$ y $c_m = c + N$. Al cupón fijo c se lo considera como un porcentaje del nominal N , es decir, $c = r_c N$ donde r_c se denomina *tasa cupón*.

Consideremos ahora un bono cupón cero con precio de mercado $P(t, T)$. Buscamos el rendimiento o “tasa de interés interna del bono”, es decir, la tasa de interés short constante que dará el mismo valor a este bono que el valor dado por el mercado. Denotando este valor de la tasa short por y , buscamos resolver la ecuación

$$P(t, T) = e^{-y \cdot (T-t)} \cdot 1,$$

donde el factor 1 indica el valor final del bono. De esta manera, tenemos la siguiente definición:

Definición 1.1.7. *Yield cupón cero continuamente compuesto.* El rendimiento o yield cupón cero continuamente compuesto, $y(t, T)$, está dado por

$$y(t, T) = -\frac{\ln P(t, T)}{T - t}.$$

Observamos que el yield $y(t, T)$ no es más que la tasa spot capitalizada continuamente para el intervalo $[t, T]$, es decir, $R(t, T)$.

Aunque los pagos de cupones y las fechas relevantes están preasignados en un bono con cupón, los cash flow futuros de poseer un bono son reinvertidos a tasas que no se conocen de antemano. Por lo tanto, el rendimiento total de un bono con cupón que se mantiene hasta el vencimiento (o por un periodo de tiempo menor) parece ser incierto.

La manera estándar para eludir esta dificultad es extender la noción de un *yield to maturity* (rendimiento a la madurez) de los bonos con cupón. Este yield to maturity se define como la tasa de interés short constante Y_c (en cada tiempo t) con la cual se descuentan todos los cupones del bono.

Definición 1.1.8. Yield to maturity. Considere un bono que paga m cupones idénticos a través de una tasa cupón r_c en las fechas $T_1 < \dots < T_m$ y el nominal N en T_m . El yield to maturity capitalizado continuamente del bono en tiempo t , denotado por $Y_c(t)$, está determinado de manera única por la siguiente relación:

$$P_C(t) = \begin{cases} \sum_{j=k}^m r_c N e^{-\tau(t, T_j) Y_c(t)} + N e^{-\tau(t, T_m) Y_c(t)}, & \text{si } T_{k-1} < t \leq T_k, \quad \forall k \geq 2; \\ \sum_{j=1}^m r_c N e^{-\tau(t, T_j) Y_c(t)} + N e^{-\tau(t, T_m) Y_c(t)}, & \text{si } T_0 \leq t \leq T_1, \end{cases}$$

donde $P_C(t)$ denota el precio del bono en el momento t tal que $T_0 \leq t \leq T_m$ y T_0 es el momento en que se emite el bono.

Cabe destacar que se podría definir la yield to maturity para bonos con cupones que varían en cada fecha de pago de los mismos. Sin embargo, no utilizaremos este tipo de bonos en el presente trabajo.

En la práctica, los yield to maturity sufren el llamado “efecto cupón”. Es decir, dos bonos con el mismo vencimiento pero con diferentes tasas cupón no necesariamente tienen el mismo yield to maturity. Para superar este efecto cupón, se suele considerar el *par yield* (rendimiento a la par). Si la tasa cupón es tal que el precio del bono en t coincide con su valor nominal N , decimos que el bono está valorado *a la par* y denominamos a la tasa cupón como el *par yield*.

Definición 1.1.9. Par Yield. El *par yield* para un bono con cupón determinado es la tasa cupón que causa que el precio del bono sea igual a su valor nominal. Luego, el *par yield* viene dado por la siguiente fórmula:

$$py_c(t) = \begin{cases} \frac{1 - D(t, T_m)}{m}, & \text{si } T_{k-1} < t \leq T_k, \quad \forall k \geq 2; \\ \sum_{j=k}^m D(t, T_j) \\ \frac{1 - D(t, T_m)}{m}, & \text{si } T_0 \leq t \leq T_1. \\ \sum_{j=1}^m D(t, T_j) \end{cases}$$

Normalmente, los par yields se utilizan en la dirección opuesta. Es decir, en lugar de utilizar los factores de descuento para resolver los par yields, a menudo se utiliza el par yield para resolver los factores de descuento. Este proceso se llama *bootstrapping*. El ejemplo clásico es el uso de tasas swap para resolver tasas cupón cero asociadas. Como veremos más adelante, las tasas swap siempre se cotizan en el mercado a la par y, como tal, reúnen todas las características de un bono valorado a la par.

Por último, diremos que un *security* es un activo o título financiero como, por ejemplo, una acción, un bono, entre otros. Luego, a cualquier *security* cuyos cash flows dependan del nivel de los precios de los bonos los llamaremos *derivados de tasa de interés*.

1.2. Estructura temporal de tasas de interés

La *estructura temporal de tasas de interés* es, sin duda, el concepto central en el estudio del mercado de renta fija. Esta estructura se puede obtener a partir de los datos del mercado de las tasas de interés y es el gráfico de la función que mapea cada madurez en la tasa de interés correspondiente en tiempo t .

En esta función se incorpora un amplio rango de información, que va desde cuestiones fundamentales tales como el valor temporal del dinero, las acciones esperadas de política monetaria y las expectativas inflacionarias a las ideas más complicadas pero igualmente importantes, como las primas de riesgo, la evaluación de la solvencia, y la liquidez relativa.

Definición 1.2.1. *Estructura temporal de tasas de interés.* La estructura temporal de tasas de interés (a la que en general nos referiremos como “curva de tasas de interés”) en tiempo t es el gráfico de la función

$$T \mapsto Y(t, T),$$

donde $Y(t, T)$ es la tasa de interés en tiempo t con madurez T .

Esta definición puede extenderse a yields de bonos además de a tasas de interés. Tal estructura temporal suele denominarse “yield curve” o “curva de rendimiento”.

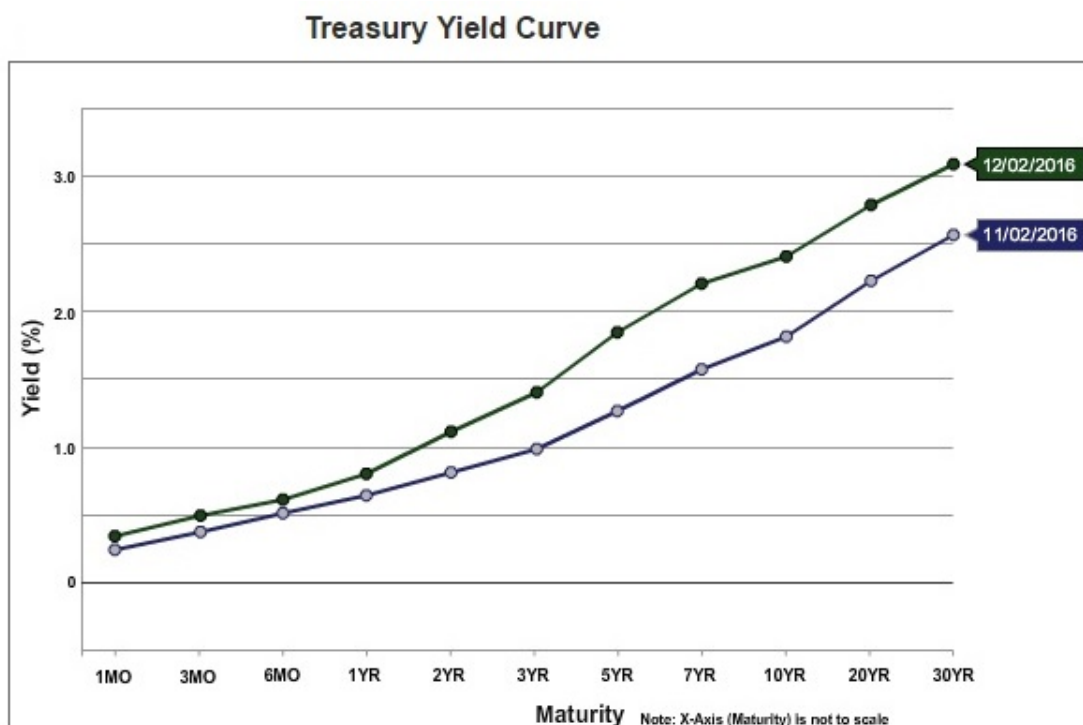
El tipo de capitalización puede variar dentro del gráfico. Por ejemplo, podríamos considerar la estructura temporal correspondiente a la siguiente función:

$$T \mapsto \begin{cases} L(t, T), & t < T \leq t + 1 \text{ (años)} \\ \tilde{Y}(t, T), & T > t + 1 \text{ (años)}. \end{cases}$$

Se trata de un gráfico en tiempo t donde $L(t, T)$ es una tasa capitalizada simplemente para todos los vencimientos T hasta un año e $\tilde{Y}(t, T)$ es una tasa de interés capitalizada anualmente con vencimientos T mayores a un año.

Un ejemplo de una estructura temporal de tasas de interés es la que observamos en la Figura 1.1.

Figura 1.1: **Curva de rendimiento del Tesoro Americano.** Fuente: U.S. Department Of The Treasury (<https://www.treasury.gov>)



No hay una única curva de tasas de interés sino una serie de diferentes curvas posibles para diferentes puntos en el tiempo, diferentes monedas, diferentes emisores, y los diferentes niveles de crédito.

En la práctica, la estructura temporal de tasas de interés se deriva de los precios de varios instrumentos de tipos de interés activamente negociados, tales como Letras del Tesoro, bonos del Tesoro, swaps y futuros. Hay que tener en cuenta que la curva en cualquier día dado está determinada exclusivamente por los precios cotizados en el mercado en ese día. A lo largo del tiempo, la curva observada puede adoptar diversas formas: puede tener pendiente positiva, ser plana, tener pendiente negativa o presentar una joroba.

La pregunta clave desde una perspectiva de manejo del riesgo es entender cómo la estructura temporal de tasas de interés evoluciona con el tiempo. El estudio histórico de la evolución de la curva de tasas de interés a través del tiempo nos permite enfatizar los cuatro siguientes importantes hechos: (1) las tasas de interés no son negativas (aunque en la actualidad se pueden encontrar casos de tasas negativas); (2) las tasas de interés tienen reversión a la media; (3) los cambios de las tasas de interés no están perfectamente correlacionados; (4) la volatilidad de las tasas a corto plazo es más alta que la volatilidad de las tasas a largo plazo. Además, se ha concluido en muchos casos que con sólo tres factores se puede explicar más del 95 % de los cambios en la curva de rendimiento. Estos factores están asociados a los cambios en el

nivel, la pendiente y la curvatura de la estructura temporal. Esta afirmación se explicará más adelante en la sección 2.2 de modelos estadísticos.

Un asunto más importante es tratar de entender estos cambios. El estudio de la estructura temporal de tasas de interés se reduce a preguntarse sobre las preferencias de los participantes del mercado (inversores, prestatarios) para los vencimientos de la curva. De hecho, si fueran indiferentes a éstos, las curvas de tasas de interés serían invariablemente planas, y la noción de estructura temporal no tendría sentido. Las preferencias de los participantes del mercado pueden ser guiadas por sus *expectativas*, la naturaleza de sus pasivos o activos y el nivel de las *primas de riesgo* que requieren para compensar su aversión al riesgo.

1.3. Swap spread

El objeto de estudio en el presente trabajo es la curva de tasas constituida por la diferencia o spread entre la curva de tasas dadas por la deuda que emite el Tesoro de Estados Unidos; específicamente bills, notas y bonos; y la curva correspondiente a las tasas que se obtienen de los contratos Swap sobre la tasa LIBOR.

1.3.1. Contratos futuros IRS: Interest Rate Swap

En los mercados internacionales actuales, como el CME (Chicago Mercantile Exchange), se negocia una gran variedad de productos financieros vinculados a tasas de interés. Entre ellos se destacan los futuros sobre la tasa Swap. Éstos se denominan Interest Rate Swap Futures o IRS Futures y consisten en el intercambio de pagos semianuales a una tasa anual fija (tasa Swap) por pagos flotantes o variables trimestrales basados en la tasa LIBOR (London InterBank Offered Rate) a 3 meses. La LIBOR es una tasa de referencia diaria basada en la tasa de interés a la cual los bancos otorgan préstamos fuera de la Reserva Federal de Estados Unidos.

El caso más común de IRS es aquel en el que se intercambia un conjunto de flujos de interés por otro, sin intercambio de nominal. Para describirlo de una manera matemática general, consideremos un IRS iniciado en el momento t_0 . Éste es un compromiso para intercambiar pagos de intereses asociados con un monto nominal N , establecidos en fechas posteriores a t_0 claramente identificadas $t_1 < t_2 < \dots < t_n$. El comprador del Swap realizará pagos fijos de tamaño $s_{t_0} \tau_i N$ y recibirá pagos flotantes de tamaño $L_{t_i} \tau_i N$ donde τ_i es $(t_{i+1} - t_i)/360$ con $i = 0, \dots, n - 1$. El vendedor del IRS es el que recibe los pagos fijos. La tasa LIBOR L_{t_i} se determina en las fechas establecidas t_i con $i = 0, 1, \dots, n - 1$. La madurez del Swap es $m = \sum_{i=0}^{n-1} \tau_i$ años y la tasa s_{t_0} es la tasa Swap en t_0 para el período $[t_0, t_n]$. Los flujos variables dependiendo de L_{t_i} son observados en tiempo t_i , pero se pagan en tiempo t_{i+1} .

La estructura de flujos del IRS se puede descomponer de la siguiente manera: para valorar el IRS solo necesitamos calcular el valor presente de cada posición usando los factores de descuento apropiados. El

valor presente de los flujos fijos, VP_{fix} , es obtenido de la siguiente manera:

$$VP_{fix} = s_{t_0}\tau_0NP(t_0, t_1) + s_{t_0}\tau_1NP(t_0, t_2) + \dots + s_{t_0}\tau_{n-1}NP(t_0, t_n). \quad (1.4)$$

donde $P(t_0, t_i)$ es el valor en tiempo t_0 de un bono cupón cero libre de riesgo que madura en t_i .

Debido a que L_{t_i} son desconocidos en t_0 , debemos usar las tasas forward conocidas $F(t_0; t_i, t_{i+1})$ con $i = 0, \dots, n - 1$ para cada período de tiempo. Luego, el valor presente del flujo variable, VP_{flo} , se calcula como

$$VP_{flo} = F(t_0; t_0, t_1)\tau_0NP(t_0, t_1) + F(t_0; t_1, t_2)\tau_1NP(t_0, t_2) + \dots \\ \dots + F(t_0; t_{n-1}, t_n)\tau_{n-1}NP(t_0, t_n) \quad (1.5)$$

Para que el IRS sea justo en tiempo t_0 éste debe valer 0 en ese instante de tiempo, es decir, ambas partes deben coincidir. Luego, igualando (1.4) y (1.5) podemos obtener el valor s_{t_0} para la tasa Swap en t_0 .

En general, utilizando la definición de tasa forward vista en el Capítulo 1, podemos expresar la tasa Swap forward en cada tiempo t a través de las tasas forward LIBOR de la siguiente manera:

Definición 1.3.1. *La tasa Swap forward $S(t; t_0, t_n)$ en tiempo $t \leq t_0$ para el conjunto de tiempos t_0, \dots, t_n y fracciones de años $\tau_0, \dots, \tau_{n-1}$ es la tasa en la parte fija del IRS que hace que este contrato sea justo en el tiempo presente, es decir, que los valores presentes de cada parte sean iguales. Luego:*

$$S(t; t_0; t_n) = \frac{1 - \prod_{j=1}^n \frac{1}{1 + \tau_{j-1} F(t; t_{j-1}, t_j)}}{\sum_{i=1}^n \tau_{i-1} \prod_{j=1}^i \frac{1}{1 + \tau_{j-1} F(t; t_{j-1}, t_j)}}.$$

1.3.2. Curva de swap spread

El diferencial o spread entre las tasas del Tesoro Estadounidense y las tasas de los futuros sobre la LIBOR o Eurodollars, que se conoce como TED spread o Swap Spread, ha sido un comercio básico en los mercados de futuros desde que el contrato Eurodollar fue incluido en 1981. Como medida del diferencial de crédito entre la deuda bancaria de alto grado y la deuda del Tesoro, el Swap Spread a menudo ha respondido fuertemente a las crisis financieras que han amenazado con dañar la calidad crediticia del sistema bancario.

El Swap Spread es un indicador del riesgo de crédito percibido en la economía general. Esto se debe a que las tasas del Tesoro se consideran libres de riesgo, mientras que la LIBOR refleja el riesgo de crédito de los préstamos a los bancos comerciales. Cuando el Swap Spread aumenta, eso es una señal de que los prestamistas creen que el riesgo de incumplimiento de préstamos interbancarios (también conocido como

riesgo de contraparte) está aumentando. Por lo tanto, los prestamistas interbancarios exigen una tasa de interés más alta, o aceptan retornos más bajos en inversiones seguras como las letras del Tesoro. Cuando se considera que el riesgo de incumplimiento bancario está disminuyendo, el Swap Spread disminuye.

Aunque el mundo se ha vuelto menos temerario a lo largo de los años y tiende a reaccionar con menos violencia a las crisis financieras de lo que alguna vez lo hizo, el Swap Spread básico sigue siendo una parte activa de la escena comercial. Y en los últimos años, el comercio se ha extendido a un territorio más sofisticado. El Swap Spread original era simplemente el diferencial entre el precio de un contrato de futuros de bills del Tesoro y el precio de un contrato futuro de Eurodólares (Eurodollars). Ahora nos encontramos con un interés cada vez mayor en el comercio de notas del Tesoro (o futuros de notas) contra strips o conjunto de futuros de Eurodólar.

Como acabamos de ver, se pueden considerar varias maneras de obtener el valor del Swap Spread. Por ejemplo, tomando un contrato de Eurodollar y un bill del Tesoro. O tomando un conjunto de contratos de Eurodollar para cubrir el plazo de una nota del Tesoro, etc.

En el presente trabajo, tomaremos los datos de estas tasas Swap Spread obtenidos de Bloomberg. Éstos se construyen de la siguiente manera:

Definición 1.3.2. *El valor de la tasa Swap Spread en tiempo t que vence en el tiempo futuro $T > t$ viene dado por:*

$$SwapSpread(t, T) = tasaSWAP(t, T) - tasaCMT(t, T);$$

donde $tasaSWAP(t, T)$ es la tasa Swap en tiempo t que vence en T y $tasaCMT(t, T)$ es la tasa Constant Maturity Treasury en t que vence en T y que describimos a continuación.

Las tasas Constant Maturity Treasury (CMT) se calculan usando los rendimientos de las bills, notas y bonos on-the-run del Tesoro de Estados Unidos (y algunos off-the-run). En los cálculos se excluyen los contratos que maduran en 1 mes, 3 meses, 6 años, 8 años, 9 años y 25 años. La metodología involucra una interpolación lineal entre dos de los genéricos que están alrededor. Por ejemplo, la tasa a 6 años se obtiene interpolando las tasas a 5 y 7 años. Además, los cálculos asumen 365 días por año y los días desde el settlement a la madurez.

La tasa Swap es la que se obtiene como describimos en la sección anterior.

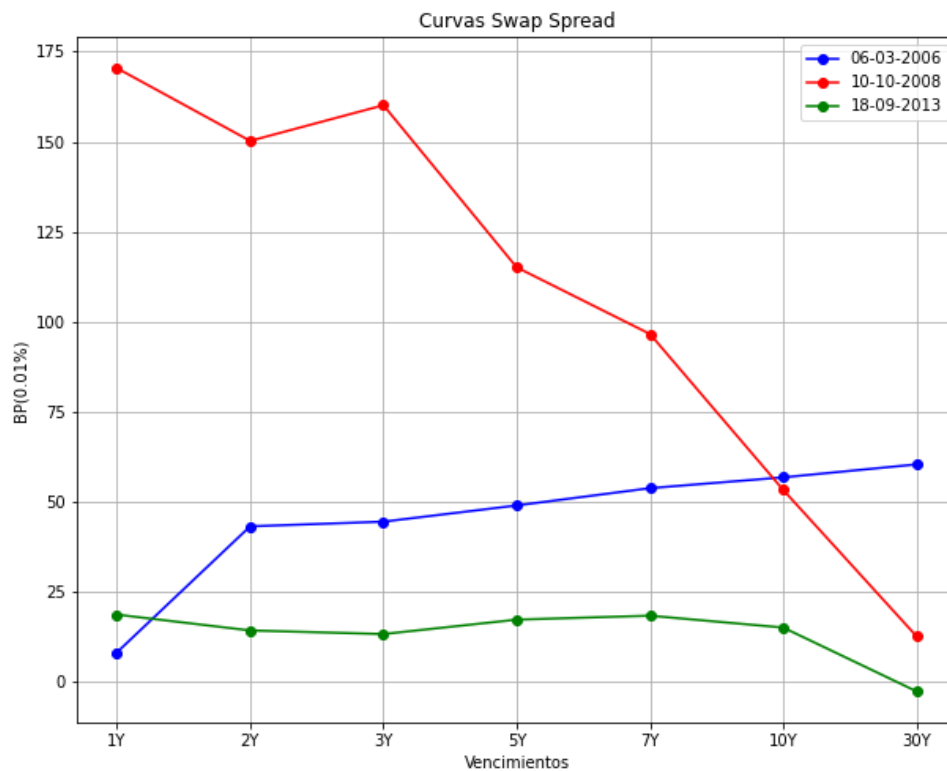
Definición 1.3.3. *La curva de tasas Swap Spread en tiempo t viene dada por el gráfico de la función*

$$T \mapsto SwapSpread(t, T)$$

donde $SwapSpread(t, T)$ es la tasa Swap Spread en tiempo t que vence en tiempo T .

Ejemplos de tal curva se puede ver en la Figura 1.2.

Figura 1.2: Curva o Estructura temporal de tasas Swap Spread. Fuente: Bloomberg. Datos facilitados por la empresa Arfima.



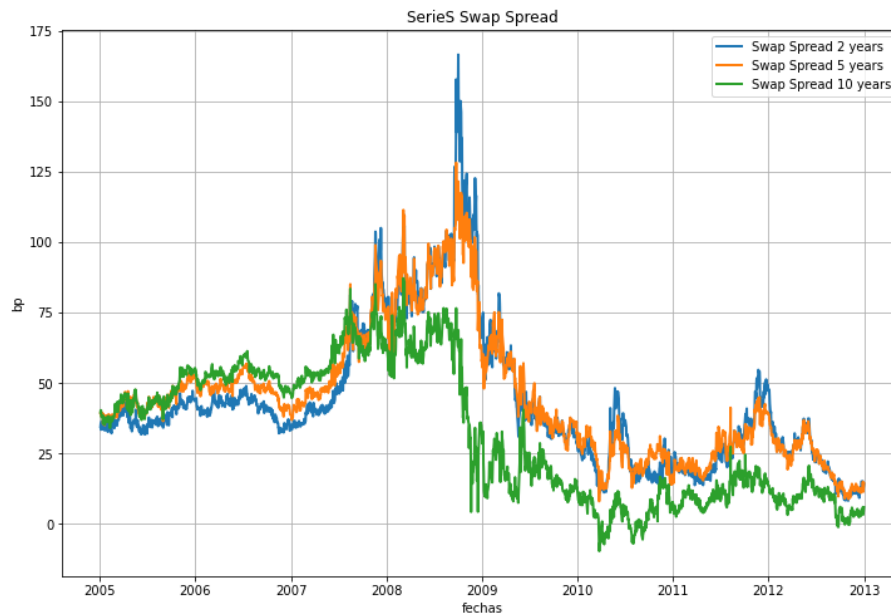
Podemos observar que la parte frontal y media de la curva aumenta cuando entramos en el período relacionado con la crisis de las hipotecas subprime.

El colapso de la burbuja inmobiliaria en Estados Unidos en el año 2006 provocó, aproximadamente en octubre de 2007, la llamada crisis de las hipotecas subprime que luego desató de manera directa la crisis financiera de 2008. Las repercusiones de la crisis hipotecaria comenzaron a manifestarse de manera extremadamente grave desde inicios de 2008, contagiándose primero el sistema financiero estadounidense, y después el internacional, teniendo como consecuencia una profunda crisis de liquidez y causando, indirectamente, otros fenómenos económicos como una crisis alimentaria global, diferentes derrumbes bursátiles (como la crisis bursátil de enero de 2008 y la crisis bursátil mundial de octubre de 2008) y, en conjunto, una crisis económica a escala internacional.

Esta crisis se ve reflejada claramente cuando observamos la serie de tiempo de las tasas Swap Spread.

En la Figura 1.3 se pueden apreciar las tasas elevadas durante el período de crisis recién mencionado.

Figura 1.3: Serie de tiempo de tasas Swap Spread. Fuente: Bloomberg. Datos facilitados por la empresa Arfima.



Capítulo 2

Modelos de Nelson y Siegel para la estructura temporal de tasas de interés

2.1. Introducción

Durante los últimos años se han hecho grandes esfuerzos tanto en el ámbito académico como profesional para intentar explicar y predecir el comportamiento de diversas curvas de tasas de interés. En consecuencia, se ha desarrollado una amplia variedad de modelos de la estructura temporal de tasas.

Tales modelos pueden clasificarse de diferentes maneras. Si se imponen restricciones a la evolución de las tasas con el fin de evitar oportunidades de ganancia libre de riesgo, entonces el modelo se conoce como libre de arbitraje. De lo contrario, se dice que el modelo es puramente estadístico. Los modelos libres de arbitraje contienen algunos ingredientes que surgen de modelos de equilibrio y por lo tanto tienen un fuerte atractivo económico. Los trabajos iniciales dentro de esta clase son Vasicek [61], Cox et al. [14] y Heath et al. [31], mientras que Nelson and Siegel [47] y Svensson [58] son trabajos pioneros en la clase de modelos estadísticos. Además, los modelos de estructura temporal pueden incluir o no directamente los factores macroeconómicos y financieros que impulsan la curva de yields. Finalmente, la relación entre factores y tasas de interés puede ser lineal o suponer una especificación más general. Ejemplos de modelos lineales son la clase de modelos afines estudiados por Duffie and Kan [24], mientras que Leippold and Wu [38] constituye un ejemplo de modelos no lineales.

En varios de los modelos mencionados se suelen encontrar deficiencias a nivel empírico. Por ejemplo, los modelos libres de arbitraje principalmente buscan ajustar la curva de tasas en un punto fijo de tiempo y, en general, no abordan cuestiones de dinámica o previsión. Entre los modelos de equilibrio, algunos se centran sólo en el ajuste dentro de la muestra y no en la predicción fuera de ella. Sin embargo, otros que sí lo hacen, no arrojan buenas predicciones (ver Duffee [22]).

Por otro lado, la estimación de tales modelos suele ser problemática. Esto se debe, principalmente, a

que en el ajuste de datos aparece una gran cantidad de estimadores de máxima verosimilitud. Esto genera una carga computacional que suele limitar la validación eficaz de modelos robustos así como de estudios de simulación completa de las propiedades de los estimadores de la muestra (ver Dai and Singleton [16]).

Un modelo parsimonioso, que produce buenas predicciones y que es flexible para adoptar diversas formas de la curva de rendimientos es el introducido por Nelson and Siegel [47]. Además, permite describir a la curva de manera estructural. En este Capítulo nos focalizaremos en describir el modelo de Nelson y Siegel en su versión estática y dinámica. Para ello, seguiremos principalmente los trabajos de Nelson and Siegel [47], Diebold and Li [18], Diebold et al. [20] y Christensen et al. [11]. Un resumen de los mismos se puede encontrar en Diebold and Rudebusch [19].

2.2. Descripción del modelo de Nelson y Siegel estático

En el trabajo de Nelson and Siegel [47], los autores describen la curva de tasas forward instantáneas con una forma funcional:

$$f(\tau) = \beta_1 + \beta_2 e^{-\lambda\tau} + \beta_3 \lambda \tau e^{-\lambda\tau}, \quad (2.1)$$

siendo $\tau = \tau(t, T) = T - t \geq 0$ el tiempo al vencimiento con t un tiempo fijo y $\lambda > 0$. Los términos no constantes de (2.1) son soluciones de una ecuación diferencial de segundo orden con raíces iguales motivada heurísticamente por la teoría de las expectativas. Este modelo se puede ver como una constante (β_1) más una función de Laguerre. Cada función de este tipo consiste en un polinomio multiplicado por un término de decaimiento exponencial y forman una clase matemática de aproximación de funciones.

Este tipo de aproximación a través de términos exponenciales garantiza, asimismo, la monotonidad y la curvatura.

De esta manera, debido a que las yields son promedios de tasas forward ($y(\tau) = \frac{1}{\tau} \int_0^\tau f(u) du$), la curva de yields manifestará el mismo rango de formas y viene dada por la siguiente función:

$$y(\tau) = \beta_1 + \beta_2 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_3 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right). \quad (2.2)$$

Notar que ésta es una representación de la sección transversal de tasas en algún punto fijo en el tiempo t y la denominaremos *Curva de Nelson y Siegel* en el tiempo t . Luego, uno puede ajustar las yields de bonos con diferentes vencimientos que se encuentran en el mercado a través de esta forma funcional para cada momento fijo de tiempo.

Esta representación satisface restricciones básicas de la teoría económica financiera:

- $P(t, t) = 1$ y $\lim_{T \rightarrow +\infty} P(t, T) = 0$, como es apropiado.
- $\lim_{T \rightarrow t} y(\tau(t, T)) = f(\tau(t, t)) = r(t)$, la tasa short instantánea y ;

- La constante β_1 asegura que en el modelo, las tasa de interés a largo plazo tiendan a esa constante, es decir, $\lim_{T \rightarrow +\infty} y(\tau(t, T)) = \beta_1$.

Además, se pueden interpretar β_1 , β_2 y β_3 como tres *factores latentes* y a los términos que acompañan a cada factor se los puede considerar como *cargas*. Luego, la carga en β_1 es 1, una constante que no tiende a cero en el límite; entonces β_1 se puede ver como un factor de largo plazo. La carga en β_2 es $(1 - e^{-\lambda\tau})/\lambda\tau$, una función con valor 1 en $\tau = 0$ pero que decrece monótonamente a cero; entonces β_2 puede verse como un factor de corto plazo. La carga en β_3 es $((1 - e^{-\lambda\tau})/\lambda\tau) - e^{-\lambda\tau}$, comienza con el valor 0 en $\tau = 0$ (y entonces no es de corto plazo), es creciente hasta alcanzar su máximo, y luego decae a cero (con lo cual no es de largo plazo); entonces β_3 se puede ver como un factor de plazo medio. Por último, cabe mencionar que el parámetro λ determina la tasa de decrecimiento exponencial en las cargas de β_2 y β_3 ; valores pequeños de λ producen un decrecimiento lento y puede ajustar mejor la curva en los plazos largos, mientras que valores grandes de λ producen un decrecimiento rápido y puede ajustar mejor la curva en plazos cortos. La constante λ también determina el punto de máximo de la carga en β_3 el cual satisface $0 < \tau < \frac{3}{\lambda}$.

Un análisis más detallado sobre las implicaciones del parámetro λ se pueden ver en el Apéndice B.

Debido a que las tasas de interés para cada vencimiento evolucionan de forma dinámica, éstas no sólo presentan una sección transversal: t fijo y τ variable, sino también una dimensión temporal: τ fijo y t variable. La situación en cuestión es, en un sentido, muy simple: modelar y predecir una serie de tiempo. Pero en otro sentido bastante más complejo e interesante es que la serie a modelar es de hecho una serie de tiempo de *curvas*. Si bien la literatura estadística está relacionada con esta cuestión (ver por ejemplo Ramsay and Silverman [49]), trata con conjuntos de curvas y no con series de tiempo de curvas. Es decir, normalmente no se ocupa de la dinámica.

2.3. Descripción del modelo de Nelson y Siegel dinámico DNS

Fijemos $T \in \mathbb{N}$ el horizonte temporal. Como comentamos en la sección anterior, la representación de Nelson y Siegel se puede interpretar de una manera dinámica como un modelo de factor latente en el que los factores de nivel (L), pendiente (S) y curvatura (C) varían en el tiempo y los términos que multiplican estos factores son las cargas de los factores. Así, podemos considerar al nivel, pendiente y curvatura como series de tiempo $\{L_t\}_{0 \leq t \leq T}$, $\{S_t\}_{0 \leq t \leq T}$ y $\{C_t\}_{0 \leq t \leq T}$ de manera tal que la dinámica de tasas en el tiempo queda expresada de la siguiente manera:

$$y(t, \tau) = L_t + S_t \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + C_t \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right), \quad \forall 0 \leq t \leq T \quad (2.3)$$

con $\tau = T - t$ y λ una constante asociada a la ecuación diferencial de la cual proviene el modelo.

Es importante mencionar que la constante λ se podría considerar como una variable que dependa del tiempo t , aunque en este trabajo asumiremos constante.

Denotemos al conjunto de tasas de interés observadas en tiempo $t \in \{1, 2, \dots, T\}$ por $\tilde{y}(t, \tau_1), \tilde{y}(t, \tau_2), \dots, \tilde{y}(t, \tau_N)$, donde $N \in \mathbb{N}_{\geq 3}$, $\tau_i = T_i - t$ y $0 < T_1 < T_2 < \dots < T_N$ los vencimientos.

Por otro lado, debido a que los valores de dicha curva en cada tiempo t provienen de observaciones, es necesario considerar el error con respecto al modelo teórico. Este error involucra errores de medición y desvíos con respecto a la curva de tasas teórica que se quiere ajustar. Es importante destacar que la diferencia con el modelo teórico se ve afectada además por el valor que se va a asumir de la constante λ . Luego, denotamos al error en cada tiempo t por $\epsilon_t(\tau_i, \lambda)$ con $i = 1, \dots, N$.

Luego, la ecuación que relaciona el conjunto de N tasas con los tres factores no observables queda expresada como:

$$\begin{pmatrix} \tilde{y}(t, \tau_1) \\ \tilde{y}(t, \tau_2) \\ \vdots \\ \tilde{y}(t, \tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{pmatrix} \begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix} + \begin{pmatrix} \epsilon_t(\tau_1) \\ \epsilon_t(\tau_2) \\ \vdots \\ \epsilon_t(\tau_N) \end{pmatrix}, \quad (2.4)$$

Ahora, para cada tiempo t , denotamos por \tilde{Y}_t , X_t , ϵ_t y Λ a los elementos de \mathbb{R}^N , \mathbb{R}^3 , \mathbb{R}^N y $\mathbb{R}^{N \times 3}$ definiendo el vector de tasas observadas en t , el vector de factores en t , el vector de errores en t y la matriz de cargas, respectivamente. Luego, tenemos:

$$\tilde{Y}_t = \begin{pmatrix} \tilde{y}(t, \tau_1) \\ \tilde{y}(t, \tau_2) \\ \vdots \\ \tilde{y}(t, \tau_N) \end{pmatrix}, \quad X_t = \begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix}, \quad \epsilon_t = \begin{pmatrix} \epsilon_t(\tau_1, \lambda) \\ \epsilon_t(\tau_2, \lambda) \\ \vdots \\ \epsilon_t(\tau_N, \lambda) \end{pmatrix}$$

y

$$\Lambda = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{pmatrix}.$$

con $t \in \{1, 2, \dots, T\}$ y $\tau_i = T_i - t$ para $i = 1, \dots, N$.

La ecuación (2.4) se puede escribir de manera reducida para cada tiempo $t \in \{1, 2, \dots, T\}$ de la siguiente manera:

$$\tilde{Y}_t = \Lambda X_t + \epsilon_t. \quad (2.5)$$

En el Apéndice B se demuestran algunas propiedades de la matriz Λ , como por ejemplo, que todas sus entradas son positivas y que tiene rango completo.

En general, se asume que el vector aleatorio ε_t tiene una distribución normal multivariada para cada $t \in \{1, 2, \dots, T\}$, y que sus componentes no están correlacionadas. Esto es, que los errores de observación de las tasas para cada vencimiento no están correlacionados entre sí. Luego,

$$\varepsilon_t \sim N(0, \Sigma(\varepsilon_t)) \quad \text{con} \quad \Sigma(\varepsilon_t) = \begin{pmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \alpha_2 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \alpha_7 \end{pmatrix}.$$

La suposición de la matriz $\Sigma(\varepsilon_t)$ diagonal, lo que implica que las desviaciones de los rendimientos de diversos vencimientos desde la curva de rendimiento no están correlacionadas, es bastante estándar. Por ejemplo, en la estimación de los modelos de estructura a término de no arbitraje, generalmente se añade “error de medida” i.i.d. a los rendimientos observados. Este supuesto también se requiere para la tratabilidad computacional dado el elevado número de rendimientos observados utilizados.

Además, tiene sentido asumir que las desviaciones o errores cometidos al observar los rendimientos en cada vencimiento sean independientes de los valores de dichos rendimientos. Es decir, que \tilde{Y}_t es independiente de ε_t o lo que es equivalente (de la suposición de \tilde{Y}_t y ε_t normales):

$$Cov[\tilde{Y}_t, \varepsilon_t] = 0.$$

Diebold and Li [18] plantean una versión convenientemente dinamizada de Nelson y Siegel como un modelo de tres factores dinámicos (DNS) de nivel, pendiente y curvatura. Específicamente, ellos interpretan de manera dinámica la representación de Nelson y Siegel como un modelo en el cual β_1 , β_2 y β_3 son factores de nivel, pendiente y curvatura (más que corto, mediano y largo plazo) variando en el tiempo. Luego, la curva de tasas tiene la siguiente representación:

$$y(t, \tau) = \beta_{1t} + \beta_{2t} \left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} \right) + \beta_{3t} \left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau} \right),$$

con $t \in [0, T]$ un instante de tiempo, $\tau = T - t \geq 0$ el tiempo al vencimiento, $\lambda_t > 0$ para todo t y donde los movimientos dinámicos de β_{1t} , β_{2t} y β_{3t} siguen un proceso AR(1) univariado. En particular, Diebold and Li [18] fijan el valor de λ_t en 0.0609 para todo t , por razones que son especificadas en su trabajo.

El modelo de Nelson y Siegel dinámico es muy popular en la práctica debido a varias razones, tanto empíricas como teóricas, como se discute en Diebold and Li [18]. Empíricamente, el modelo DNS es estable y simple para estimar. Además, es muy flexible y se ajusta notablemente bien tanto a la sección transversal como a las series de tiempo de los rendimientos en muchos países y períodos, y para muchos

tipos de bonos. Teóricamente, este modelo impone ciertas propiedades económicamente deseables, como se mostró previamente para el caso estático.

La literatura contiene una gran variedad de métodos para construir factores de tasas de bonos y cargas de factores. Tanto el modelo de Nelson y Siegel estático como el modelo de Nelson y Siegel dinámico sitúan a la estructura sólo en las cargas, dejando a los factores libres.

Otro enfoque coloca a la estructura sólo en los factores estimados, dejando a las cargas libres. Por ejemplo, utilizando un análisis de componentes principales, los factores podrían ser los primeros componentes principales. Éstos se limitan a ser mutuamente ortogonales, mientras que las cargas se dejan sin restricciones. Por ejemplo, Litterman and Sheinkman [40] construyen un modelo de tres factores *latentes* a partir de un análisis de componentes principales donde concluyen que hay tres influencias comunes principales en la variación de retornos de bonos. Al primer factor lo denominan *factor de nivel*, pues representa esencialmente cambios paralelos en las tasas. Al segundo factor lo denominan *factor de pendiente* ya que cambios en este factor producen un decrecimiento en tasas de bonos a corto plazo pero aumentan las tasas de plazos largos. Por último, al tercer factor lo llaman *factor de curvatura* debido a que las cargas presentan una forma de joroba en los plazos intermedios.

Un tercer enfoque, el modelo de factores latente dinámico de no arbitraje, que es el modelo de elección en las finanzas, restringe ambos: factores y cargas factoriales. La subclase más común de este tipo de modelos es la mencionada en una sección previa y es la referida a modelos afines en la tradición de Duffie and Kan [24]. En este marco, se postulan formas lineales o afines para los factores latentes y se derivan las restricciones asociadas a las cargas factoriales para que se asegure la ausencia de arbitraje.

Sin embargo, como mencionamos al comienzo de esta sección, la estimación de tales modelos suele ser problemática a pesar de la rigurosidad evidente. En vistas de esto, Christensen et al. [11] desarrollaron una nueva clase de modelos afines de no arbitraje sobre la base de la representación de la curva de rendimiento introducida por Nelson and Siegel [47] y extendida a entornos dinámicos por Diebold and Li [18]. Si bien DNS tiene ventajas tanto empíricas como teóricas, este modelo falla en un aspecto teórico importante: no impone las restricciones necesarias para eliminar oportunidades de arbitraje. Esto motivó a Christensen et al. [11] a introducir la clase de modelos de Nelson y Siegel libre de arbitraje (AFNS) que son modelos de estructura a plazo afines que mantienen la estructura de factor de carga de DNS.

Estos autores analizan el modelo AFNS teóricamente, relacionándolo con la representación canónica (Dai and Singleton [15]) de modelos afines libres de arbitraje de tres factores. Además, el modelo AFNS arroja beneficios a nivel empírico pues las restricciones de los parámetros de Nelson y Siegel facilitan enormemente la estimación, lo que permite escapar del entorno de estimación desafiante de los modelos de tipo afín de tres factores (ver Singleton [55], Cap. 12). El entorno de estimación del AFNS es simple y robusto, y los datos favorecen fuertemente la especificación de dicho modelo cuando se considera el caso

particular en que las variables de estado pueden interactuar dinámicamente y/o sus shocks pueden estar correlacionados (correlated-factor AFNS model).

En cuanto a la predicción fuera de la muestra, se demostró una mejoría para los modelos más parsimoniosos cuando se considera el caso que las variables de estados son independientes bajo la medida de probabilidad real (independent-factor AFNS model). Los resultados también sugieren que los beneficios se pueden lograr mediante la imposición de ausencia de arbitraje, en particular para vencimientos de tasas y pronóstico de horizontes a mediano y largo plazo, aunque la evidencia es mucho menos concluyente que para un ajuste dentro de la muestra. Para el modelo AFNS independent-factor se observa una buena predicción fuera de muestra.

Finalizamos esta sección observando que los modelos hasta aquí expuestos, ya sean de factores latentes u observables, no presentan relación explícita a variables macroeconómicas. Los modelos que identifican los factores con, por ejemplo: el nivel de la curva, la pendiente o inclinación, etc, sólo logran representar efectos observados en la curva de tasas más que describir las fuentes económicas que afectan a la misma.

2.3.1. Modelos Macro-Financieros

Los diversos modelos existentes que intentan explicar la estructura a término de las tasas de interés son motivados por diferentes objetivos o demandas, por ejemplo, la previsión o simulación de tasas de interés, la valoración de bonos u opciones, o la supervisión del mercado. Aún así, existe una enorme brecha entre los modelos de la curva de rendimiento desarrollados por los macroeconomistas, que se centran en el papel de las expectativas de inflación y la actividad económica real futura en la determinación de los rendimientos, y los modelos empleados por los economistas financieros, que evitan cualquier rol explícito para tales determinantes.

Como notamos anteriormente, en el modelo de finanzas canónico, la tasa de interés a corto plazo es una función lineal simple de algunos factores no observados a veces denominados “nivel, pendiente y curvatura” pero sin interpretación económica. Las tasas de interés a largo plazo están relacionadas a esos mismos factores y los movimientos en los rendimientos a largo plazo se determinan en mayor medida por los cambios en las primas de riesgo, que también dependen de los factores latentes. En contraste, en la literatura macro, la tasa de interés de corto plazo se fija por el banco central de acuerdo a los objetivos de estabilización macroeconómica. Por ejemplo, la tasa short se puede determinar en función de las desviaciones de la inflación y la salida de los objetivos fijados por el banco central. Por otra parte, la literatura macro comúnmente ve los rendimientos a largo plazo como determinados ampliamente por las expectativas de las tasas de interés a corto plazo futuras, que a su vez dependen de las expectativas de las variables macro; es decir, los posibles cambios en las primas de riesgo a menudo se ignoran y se emplea

la hipótesis de las expectativas de la estructura temporal.

Una perspectiva macro-financiera conjunta sugeriría que comprender la manera en que los bancos centrales mueven la tasa de corto plazo en respuesta a los shocks macroeconómicos fundamentales debe explicar los movimientos en la parte corta de la curva de rendimiento; además, con la consistencia entre las tasas de corto y largo plazo impuestas por la suposición de no arbitraje, la variación futura macroeconómica esperada debe dar cuenta de los movimientos más alejados en la curva de rendimiento. Uno de los desafíos es reducir este desfase mediante la formulación y la estimación de un modelo de curva de rendimiento que integre factores macroeconómicos y financieros.

En esta dirección avanzaron Diebold et al. [20], quienes analizaron el uso de un modelo de factor latente para la curva de tasas pero incorporando explícitamente factores macroeconómicos. Particularmente, examinaron las correlaciones entre los factores de tasas en Nelson y Siegel y las variables macroeconómicas, y encontraron que el factor de nivel está altamente correlacionado con la inflación y que el factor pendiente está altamente correlacionado con la actividad real. El factor curvatura no aparece relacionado a ninguna de las principales variables macroeconómicas.

En el marco de modelos de tipo afín de no arbitraje, el papel de las variables macroeconómicas es explorado en varios artículos. En Ang and Piazzesi [2], los factores macroeconómicos son medidas de inflación y actividad real y la dinámica conjunta de los factores macro y los factores latentes adicionales son capturados por el vector de autorregresión. Para evitar depender de una serie macro específica, ellos construyen sus medidas de la actividad real y la inflación como el primer componente principal de un gran conjunto de candidatos de series macroeconómicas. Además, encuentran que los shocks de salida tienen un impacto significativo en los rendimientos intermedios y la curvatura, mientras que las sorpresas de inflación tienen grandes efectos sobre el nivel de toda la curva de rendimientos. Para una mejor tratabilidad en la estimación, Ang and Piazzesi [2] sólo permiten la dinámica unidireccional en su modelo libre de arbitraje; en concreto, las variables macro ayudan a determinar los rendimientos, pero no a la inversa. En contraste, Diebold et al. [20] consideran una caracterización bidireccional de las interacciones dinámicas macro/tasas. Ellos encuentran que la causalidad desde la macroeconomía a las tasas es de hecho significativamente más fuerte que en la dirección inversa, pero que las interacciones en ambas direcciones pueden ser importantes. Ang et al. [3] también permiten vínculos macro-financieros bidireccionales, pero adhiriendo la restricción de no arbitraje, lo que plantea un serio desafío de estimación. Ellos encuentran que la cantidad de variación en los rendimientos que se puede atribuir a factores macro depende de si el sistema permite vínculos bidireccionales. Cuando las interacciones se restringen a ser unidireccionales (de macro a factores de tasas), los factores macro sólo pueden explicar una pequeña porción de la varianza de los rendimientos de largo plazo. En contraste, cuando las interacciones pueden ser bidireccionales, el sistema atribuye más de la mitad de la varianza de los rendimientos a largo plazo a factores macro.

Resultados similares en un entorno más robusto se reportan en Bibkov and Chernov [4].

Por otra parte, Rudebusch and Wu [52] proporcionan un ejemplo de una especificación macro-financiera que emplea una estructura más macroeconómica e incluye expectativas racionales y elementos inerciales. Obtienen un buen ajuste a los datos con un modelo que combina una especificación dinámica de no arbitraje afín para los rendimientos y un pequeño modelo macro bastante estándar que consiste en una función de reacción de política monetaria, una ecuación de Euler de salida, y una ecuación de inflación.

Finalizando, en Rudebusch [50] se realiza un estudio que considera tres aspectos recientes de la investigación macro-financiera centrada en los vínculos entre las tasas de interés y la economía. En el primero se analizan los modelos que agregan variables y estructura macroeconómica a una representación financiera canónica libre de arbitraje de la curva de rendimientos. En particular, se describe en detalle el modelo de Rudebusch and Wu [52] previamente mencionado el cual integra los factores de rendimiento dentro de una estructura macroeconómica. Este análisis combina un modelo de estructura a término libre de arbitraje afín con un pequeño modelo macroeconómico de expectativas racionales y con la tasa de interés short relacionada a variables macroeconómicas a través de una función de reacción de política monetaria. Esta estructura adicional facilita la interpretación de una retroalimentación bidireccional entre los factores de estructura a plazo y las variables macro.

La segunda línea de investigación se centra en las implicaciones financieras de la valoración de bonos en un modelo macroeconómico estándar: el modelo de equilibrio general macroeconómico estocástico y dinámico ó modelo DSGE macroeconómico. Como una cuestión teórica, los precios de los activos y la macroeconomía están ligados de una manera complicada. Sin embargo, la importancia de modelar conjuntamente ambas variables macroeconómicas y los precios de los activos dentro de un marco DSGE recién ha comenzado a apreciarse. Desafortunadamente, el marco estándar del modelo DSGE parece lamentablemente inadecuado para valorar precios de los bonos aunque hay algunas modificaciones de dicho modelo que prometen mejores resultados. De hecho, en este trabajo, se introduce una versión alternativa para hacer coincidir momentos macroeconómicos y financieros en los datos.

El tercer enfoque apunta a las dificultades computacionales prácticas que aparecen en la estimación de los modelos libres de arbitraje afines y que han complicado enormemente su extensión en aplicaciones macro-financieras. En este trabajo, Rudebusch se centra en una nueva clase de modelos, los modelos libre de arbitraje Nelson y Siegel (AFNS) de la estructura a término (ver Christensen et al. [11]), que son empíricamente manejables y proporcionan un marco bastante útil para examinar diversas cuestiones macro-financieras.

Capítulo 3

Cointegración y modelo de corrección de error

En el siguiente Capítulo desarrollaremos los conceptos matemáticos de integración y cointegración y abordaremos el Modelo de Corrección de Error al que nos referiremos como VECM por las siglas de su denominación en inglés. Para ello, seguiremos principalmente el libro de Lütkepohl [43] comenzando con los conceptos básicos en los que se sustenta dicha teoría. También nos apoyaremos en los libros de Tsay [59], Lütkepohl and Krätzig [44] y Johansen [35].

3.1. Procesos autorregresivos y condición de estabilidad

El objeto de interés en lo que vamos a desarrollar es el siguiente modelo VAR(p) o modelo autorregresivo vectorial de orden p ,

$$y_t = \nu + A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t, \quad t = 0, \pm 1, \pm 2, \dots \quad (3.1)$$

donde $y_t = (y_{1t}, \dots, y_{Kt})$ es un vector aleatorio $K \times 1$, las matrices de coeficientes A_i son fijas y de dimensión $K \times K$, $\nu = (\nu_1, \dots, \nu_k)$ es un vector fijo de términos de interceptos permitiendo la posibilidad de una media distinta de cero $E(y_t)$. Finalmente, $u_t = (u_{1t}, \dots, u_{kt})$ es un *proceso de innovación o ruido blanco* K -dimensional, es decir, $E(u_t) = 0$, $E(u_t u_t') = \Sigma_u$ y $E(u_t u_s') = 0$ para $s \neq t$. La notación A' denota la transpuesta de la matriz A . Es importante destacar que la matriz de covarianza Σ_u se supone no singular si no se indica lo contrario.

3.1.1. Procesos VAR(1)

Para investigar las implicaciones del modelo, consideremos el modelo VAR(1)

$$y_t = \nu + A_1 y_{t-1} + u_t. \quad (3.2)$$

Si este mecanismo de generación comienza en algún momento, por ejemplo $t = 1$, obtenemos

$$\begin{aligned} y_1 &= \nu + A_1 y_0 + u_1, \\ y_2 &= \nu + A_1 y_1 + u_2 = \nu + A_1(\nu + A_1 y_0 + u_1) + u_2 \\ &= (I_K + A_1)\nu + A_1^2 y_0 + A_1 u_1 + u_2, \\ &\vdots \\ y_t &= (I_K + A_1 + \cdots + A_1^{t-1})\nu + A_1^t y_0 + \sum_{i=0}^{t-1} A_1^i u_{t-i} \\ &\vdots \end{aligned} \quad (3.3)$$

donde I_K es la matriz identidad de dimensión $K \times K$. Por lo tanto, los vectores y_1, \dots, y_t , están determinados de manera única por y_0, u_1, \dots, u_t . Además, la distribución conjunta de y_1, \dots, y_t está determinada por la distribución conjunta de y_0, u_1, \dots, u_t .

Aunque a veces suponemos que un proceso se inicia en un período específico, en ciertos casos es conveniente suponer que se ha iniciado en el pasado infinito. Esta suposición se hace de hecho en (3.1). ¿Qué tipo de proceso es consistente con el mecanismo (3.1) en ese caso? Para investigar esta cuestión, consideramos nuevamente el proceso VAR(1) en (3.2). De (3.3) tenemos que

$$\begin{aligned} y_t &= \nu + A_1 y_{t-1} + u_t \\ &= (I_K + A_1 + \cdots + A_1^j)\nu + A_1^{j+1} y_{t-j-1} + \sum_{i=0}^j A_1^i u_{t-i}. \end{aligned}$$

Si todos los autovalores de A_1 tienen módulo menor que 1, la sucesión A_1^i , $i = 0, 1, \dots$ es absolutamente sumable (ver Apéndice A, Sección A.9.1 del libro Lütkepohl [43]). Entonces, la suma infinita

$$\sum_{i=1}^{\infty} A_1^i u_{t-i}$$

existe en media cuadrática (ver Apéndice C, Proposición C.9 del libro Lütkepohl [43]). De esta manera,

$$\lim_{j \rightarrow \infty} (I_K + A_1 + \cdots + A_1^j)\nu = (I_K - A_1)^{-1}\nu,$$

(ver Apéndice A, Sección A.9.1 del libro Lütkepohl [43]). Además, A_1^{j+1} converge a cero exponencialmente cuando $j \rightarrow \infty$ y, por lo tanto, ignoramos el término $A_1^{j+1}y_{t-j-1}$ en el límite. Luego, si todos los autovalores de A_1 tienen módulo menor que 1, al decir que y_t es el proceso VAR(1) dado por (3.2), queremos decir que y_t es el proceso estocástico bien definido

$$y_t = \mu + \sum_{i=0}^{\infty} A_1^i u_{t-i}, \quad t = 0, \pm 1, \pm 2, \dots \quad (3.4)$$

donde

$$\mu \doteq (I_K - A_1)^{-1}\nu.$$

Como observamos anteriormente, las distribuciones y distribuciones conjuntas de los y_t están determinadas de manera única por las distribuciones del proceso u_t . Del Apéndice C.3, Proposición C.10 del libro de Lütkepohl [43], se ve que el primer y el segundo momento del proceso y_t vienen dados por

$$E(y_t) = \mu, \quad \text{para todo } t$$

y

$$\begin{aligned} \Gamma_y(h) &\doteq E((y_t - \mu)(y_{t-h} - \mu)') \\ &= \lim_{n \rightarrow \infty} \sum_{i=0}^n \sum_{j=0}^n A_1^i E(u_{t-i} u_{t-h-j}') (A_1^j)' \\ &= \lim_{n \rightarrow \infty} \sum_{i=0}^n A_1^{h+i} \Sigma_h (A_1^i)' \\ &= \sum_{i=0}^{\infty} A_1^{h+i} \Sigma_h (A_1^i)' \end{aligned}$$

debido a que $E(u_t u_s') = 0$ para $s \neq t$ y $E(u_t u_t') = \Sigma_u$ para todo t .

Como la condición para los autovalores de la matriz A_1 es importante, damos la siguiente definición:

Definición 3.1.1. *Llamamos a un proceso VAR(1) estable si todos los autovalores de A_1 tienen módulo menor que 1. Esta condición es equivalente a*

$$\det(I_K - A_1 z) \neq 0 \quad \text{para } |z| \leq 1 \quad (3.5)$$

La última condición se obtiene a partir de la Regla (7) del Apéndice A.6 de Lütkepohl [43].

Es importante señalar que el proceso y_t para $t = 0, \pm 1, \pm 2, \dots$ puede estar bien definido aún si la condición de estabilidad (3.5) no se cumple. No lo haremos aquí porque siempre asumiremos estabilidad en los procesos definidos para todo $t \in \mathbb{Z}$.

En general, podemos describir al proceso y_t definido en (3.1) a través de un operador de retardo L que definimos de la siguiente manera:

Definición 3.1.2. Sea y_t un proceso VAR(p). El operador de retardo L se define tal que

$$Ly_t \doteq y_{t-1}$$

El operador de retardo “retrocede” el índice en un período. Debido a esta propiedad, a veces se lo llama *operador de cambio, de retroceso, lag o backshift*.

De este operador derivamos la siguiente definición:

Definición 3.1.3. Sea y_t un proceso VAR(p). La diferencia de orden $d \geq 0$ del proceso y_t se define como

$$\Delta^d y_t \doteq (I_K - L)^d y_t.$$

En particular, $\Delta^0 y_t = y_t$

Luego, (3.1) puede escribirse como

$$\begin{aligned} y_t &= \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t \\ &= \nu + (A_1 L + \cdots + A_p L^p) y_t + u_t \end{aligned}$$

o

$$A(L)y_t = \nu + u_t,$$

donde

$$A(L) \doteq I_K - A_1 L - \cdots - A_p L^p. \quad (3.6)$$

Se puede ver que si $\det(A(z)) \neq 0$ para todo z tal que $|z| \leq 1$, existe un polinomio Φ tal que $\Phi(z) = \sum_{i=0}^{\infty} \phi_i z^i$ con coeficientes matriciales ϕ_i absolutamente sumables y tal que $\Phi(L)A(L) = I_K$. Luego, por proposición C.9 del Apéndice C.3 de Lütkepohl [43], el proceso $\Phi(L)u_t$ está bien definido e

$$\begin{aligned} y_t &= \Phi(L)A(L)y_t \\ &= \Phi(L)\nu + \Phi(L)u_t \\ &= \sum_{i=0}^{\infty} \phi_i L^i \nu + \sum_{i=0}^{\infty} \phi_i L^i u_t \\ &= \left(\sum_{i=0}^{\infty} \phi_i \right) \nu + \sum_{i=0}^{\infty} \phi_i u_{t-i}, \end{aligned}$$

también está bien definido.

A continuación damos la definición de operador inversible:

Definición 3.1.4. Sea y_t un VAR(p) como en (3.1) y $A(L)$ el operador (3.6). Decimos que $A(L)$ es un operador inversible si $\det(A(z)) \neq 0$ para todo z tal que $|z| \leq 1$. Además, llamamos inverso de $A(L)$ y lo denotamos por $A^{-1}(L)$ al operador $\Phi(L)$ tal que

$$\Phi(L)A(L) = I_K.$$

3.1.2. Procesos VAR(p) y la condición de estabilidad.

La discusión previa puede extenderse fácilmente a procesos VAR(p) con $p > 1$ porque cualquier proceso VAR(p) se puede escribir en forma VAR(1). Más precisamente, si y_t es un proceso VAR(p) como en (3.1), un VAR(1) Kp -dimensional correspondiente se puede definir como

$$Y_t = \boldsymbol{\nu} + \mathbf{A}Y_{t-1} + U_t,$$

donde

$$Y_t = \begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{pmatrix}, \quad \boldsymbol{\nu} = \begin{pmatrix} \nu \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

$$\mathbf{A} = \begin{pmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_K & 0 & \cdots & 0 & 0 \\ 0 & I_K & \cdots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & I_K & 0 \end{pmatrix} \quad U_t = \begin{pmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

con $Y_t \in \mathbb{R}^{Kp \times 1}$, $\boldsymbol{\nu} \in \mathbb{R}^{Kp \times 1}$, $\mathbf{A} \in \mathbb{R}^{Kp \times Kp}$ y $U_t \in \mathbb{R}^{Kp \times 1}$ para cada t .

Luego, por lo visto anteriormente para el caso $p = 1$, tenemos que Y_t es estable si

$$\det(I_{Kp} - \mathbf{A}z) \neq 0 \quad \text{para } |z| \leq 1. \quad (3.7)$$

Su vector de media es

$$\boldsymbol{\mu} \doteq E(Y_t) = (I_{Kp} - \mathbf{A})^{-1}\boldsymbol{\nu} \quad (3.8)$$

y las autocovarianzas son

$$\Gamma_Y(h) = \sum_{i=0}^{\infty} \mathbf{A}^{h+i} \Sigma_U (\mathbf{A}^i)' \quad (3.9)$$

donde $\Sigma_U \doteq E(U_t U_t')$. Usando la matriz $K \times Kp$

$$J \doteq [I_K : 0 : \cdots : 0],$$

el proceso y_t se obtiene como $y_t = JY_t$.

Debido a que Y_t es un proceso estocástico bien definido, lo mismo es cierto para y_t . Su media es $E(y_t) = J\boldsymbol{\mu}$ la cual es constante para todo t y las autocovarianzas $\Gamma_y(h) = J\Gamma_Y(h)J'$ son también invariantes en el tiempo.

Se puede ver que

$$\det(I_{Kp} - \mathbf{A}z) = \det(I_K - A_1z - \cdots - A_pz^p).$$

Dada la definición del polinomio característico de una matriz, llamamos a este polinomio el *polinomio característico inverso* del proceso VAR(p). Entonces, el proceso (3.1) es *estable* si su polinomio característico inverso no tiene raíces en el círculo unitario complejo.

Un caso de gran interés se presenta cuando una de estas raíces es igual a 1. Por ejemplo, si un proceso AR(1) tiene una raíz en $z = 1$, dicho proceso es una *caminata aleatoria*. Si bien desarrollaremos este concepto en la siguiente sección, a continuación damos la siguiente definición:

Definición 3.1.5. *Un proceso VAR(p) de dimensión K tiene una raíz unitaria si su polinomio característico inverso $p(z) = \det(I_K - A_1z - \cdots - A_pz^p)$ tiene una raíz en $z = 1$ y las demás raíces fuera del círculo unitario complejo.*

En general, decimos que un proceso tiene d raíces unitarias si tiene una raíz unitaria con multiplicidad d .

Volviendo al concepto de estabilidad formalizamos éste de la siguiente manera:

Definición 3.1.6. *El proceso VAR(p) y_t es estable si*

$$\det(I_K - A_1z - \cdots - A_pz^p) \neq 0 \quad \text{para } |z| \leq 1. \quad (3.10)$$

Esta condición se denomina condición de estabilidad.

Equivalentemente, y_t es un proceso VAR(p) estable si (3.10) ocurre e

$$y_t = JY_t = J\boldsymbol{\mu} + J \sum_{i=0}^{\infty} \mathbf{A}^i U_{t-i}.$$

Debido a que $U_t \doteq (u_t, 0, \dots, 0)$ involucra el proceso de ruido blanco u_t , el proceso y_t queda determinado por su ruido blanco o proceso de innovación. En ciertos casos, se hacen suposiciones específicas con respecto a u_t que determinan el proceso y_t por la convención anterior. Un ejemplo importante es la suposición de que u_t es *ruido blanco Gaussiano*, es decir, $u_t \sim N(0, \Sigma_u)$ para todos los t y u_t y u_s son independientes para $s \neq t$. En ese caso, se puede demostrar que y_t es un proceso gaussiano,

es decir, (y_t, \dots, y_{t+h}) tiene distribución normal multivariada para todo t y todo h .

La última ecuación también puede escribirse de la siguiente manera:

$$\begin{aligned}
 y_t &= JY_t = J\mu + J \sum_{j=0}^{\infty} \mathbf{A}^j U_{t-j} \\
 &= J\mu + \sum_{j=0}^{\infty} J\mathbf{A}^j J' J U_{t-j} \\
 &= \mu + \sum_{j=0}^{\infty} \phi_j u_{t-j}.
 \end{aligned} \tag{3.11}$$

Aquí $\mu \doteq J\mu$, $\phi_j \doteq J\mathbf{A}^j J'$ y, debido a la estructura especial del proceso de ruido blanco U_t , tenemos que $U_t = J' J U_t$ y $J U_t = u_t$. Debido a que los A_j son absolutamente sumables, lo mismo es cierto para los ϕ_j .

Nos referiremos a la representación (3.11) de y_t como la *representación de predicción de error* o *representación fundamental o canónica de media móvil*.

3.1.3. Procesos estacionarios

Un proceso estocástico y_t se dice *estacionario* si su primer y segundo momento no dependen de t . Específicamente,

Definición 3.1.7. *Un proceso estocástico y_t es estacionario si*

$$E(y_t) = \mu \quad \text{para todo } t \tag{3.12}$$

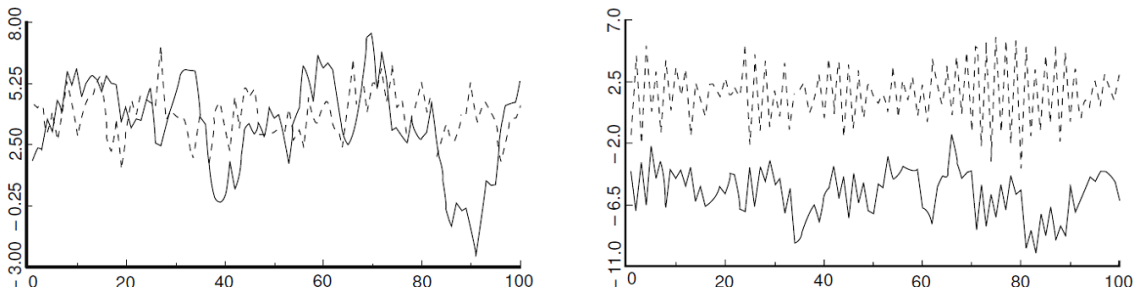
y

$$E((y_t - \mu)(y_{t-h} - \mu)') = \Gamma_y(h) = \Gamma_y(-h)' \quad \text{para todo } t \text{ y } h = 0, 1, 2, \dots \tag{3.13}$$

La condición (3.12) significa que todos los y_t tienen el mismo vector de media finito constante μ . Por lo tanto, una serie de tiempo generada por un proceso estocástico estacionario debe fluctuar alrededor de una media constante y no tener una tendencia, por ejemplo. La segunda condición (3.13) requiere que las autocovarianzas del proceso no dependan de t sino del período de tiempo h en que los dos vectores y_t y y_{t-h} están separados.

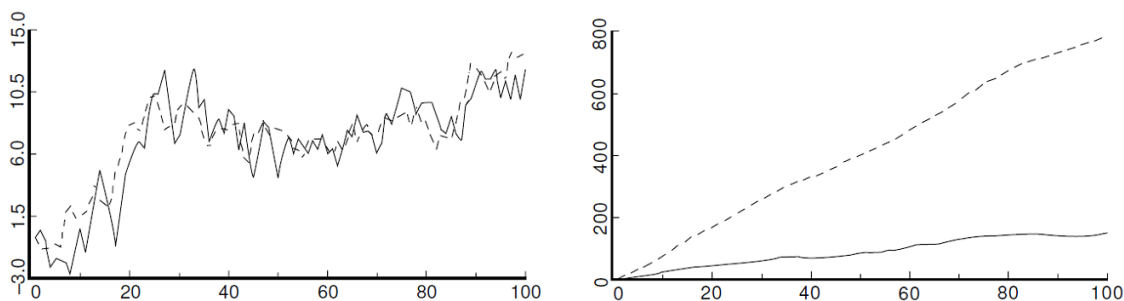
En la Figura 3.1.3 podemos observar ejemplos de series de tiempo bivariadas generadas por procesos VAR(2) estacionarios:

Figura 3.1: Series de tiempo bivariadas generadas por procesos estacionarios.



En la siguiente figura podemos observar ejemplos de series de tiempo generadas por procesos no estacionarios:

Figura 3.2: Series de tiempo bivariadas generadas por procesos no estacionarios.



Notar que, si no se indica lo contrario, se supone que todas las cantidades son finitas. Por ejemplo, μ es un vector de media con elementos finitos y $\Gamma_y(h)$ es una matriz de covarianzas finitas.

Según la primera definición, el proceso de ruido blanco u_t usado en (3.1) es un ejemplo obvio de un proceso estacionario. Además, utilizando la expresión de media móvil de y_t , sabemos que un proceso VAR(p) estable es estacionario. Establecemos este hecho como una proposición.

Proposición 3.1.8. (Condición de estacionariedad) *Un proceso y_t que es VAR(p) y estable con $t = 0, \pm 1, \pm 2, \dots$, es estacionario.*

Demostración: Se deduce de la expresión de un VAR(p) dada por (3.11) y de las propiedades de los u_t y los ϕ_i (ver sección 2.1.2 de Lütkepohl [43]).

Debido a que estabilidad implica estacionariedad, la condición de estabilidad (3.10) comúnmente se denomina condición de estacionariedad en la literatura de series de tiempo. La inversa de la Proposición

no es verdad. En otras palabras, un proceso inestable no es necesariamente no estacionario. Debido a que los procesos estacionarios inestables no son de interés en este trabajo, no discutiremos esta posibilidad.

Observación: Si y_t es un VAR estable se puede elegir la variable inicial de modo que sea estacionario incluso si el proceso comienza en algún período dado. Pero la inversa no es cierta. Luego, es posible que un proceso no estable sea estacionario, es decir, con raíces en la circunferencia unidad ($|z| = 1$) o dentro del círculo ($|z| < 1$). Por ejemplo, el proceso $y_t \equiv 0$ es estacionario, pero también responde al modelo donde los A_i pueden tomar cualquier valor. Luego, se pueden escoger aquellos A_i tales que el polinomio característico tenga raíces en el círculo, obteniendo un VAR no estable pero estacionario ($y_t \equiv 0$ implica media y varianza cero).

Antes de proseguir con nuestro análisis sobre los procesos estacionarios, damos otras variantes sobre conceptos de estacionariedad que a veces se utilizan en la literatura.

Una de ellas surge de suponer que la distribución conjunta de n vectores consecutivos cualesquiera $y_t, y_{t+1}, \dots, y_{t+n-1}$ no depende de t . Llamamos a este tipo de proceso *estrictamente estacionario*. También, podríamos tener que considerar un proceso que comience en un tiempo fijo (por ejemplo, en $t = 0$) y entonces es posible que necesite algún período de inicio hasta que los momentos se estabilicen. De hecho, es concebible que los momentos lleguen a un estado constante solo de forma asintótica. Esto sucede con frecuencia si el proceso puede convertirse en estacionario modificando los términos iniciales del proceso. En ese caso, el proceso se dice *asintóticamente estacionario*. No siempre distinguiremos entre estacionariedad asintótica y estacionariedad, pero llamaremos a un proceso estacionario si se puede lograr la estacionariedad modificando algunas variables iniciales.

3.1.4. Procesos estacionarios con tendencia

Un proceso se denomina *trend-stationary* o *estacionario con tendencia* si está representado a través del siguiente modelo:

$$y_t = \mu_0 + \mu_1 t + u_t,$$

donde u_t es una serie de tiempo estacionaria y μ_0 y μ_1 son parámetros constantes. Este modelo exhibe tendencia lineal en la media. De hecho, $E(y_t) = \tilde{\mu}_0 + \mu_1 t$ con $\tilde{\mu}_0 = \mu_0 + E(u_t)$, pero la varianza no depende de t . Esto es claro del hecho que $Var(y_t) = Var(u_t)$. Luego, las series *trend-stationary* se pueden transformar en una serie estacionaria removiendo la tendencia temporal a través de un análisis de regresión lineal simple. Este es un caso particular de un proceso con *tendencia determinística* el cual se caracteriza por el hecho de que las desviaciones desde la tendencia (dada por la media) no contribuyen al desarrollo a largo plazo del proceso. Formalmente,

Definición 3.1.9. Un proceso y_t se dice que exhibe tendencia determinística si

$$E(y_t) = d(t) \quad y \quad Var(y_t) = \sigma,$$

con d una función determinística de t no constante y σ un parámetro fijo.

Por otro lado, existen procesos donde el componente aleatorio afecta el comportamiento a largo plazo de dicho proceso, por ejemplo, el término u_t en el proceso VAR(1) definido en (3.2). Este es el caso de los procesos con tendencia estocástica que definimos a continuación:

Definición 3.1.10. Un proceso y_t se dice que exhibe tendencia estocástica si

$$E(y_t) = d(t) \quad y \quad Var(y_t) = v(t),$$

con d y v funciones determinísticas de t , ambas no constantes.

3.2. Procesos integrados

Recordemos que un proceso VAR(p),

$$y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t, \quad (3.14)$$

es estable si el polinomio definido por

$$\det(I_K - A_1 z - \dots - A_p z^p)$$

no tiene raíces en el borde y dentro del círculo unitario complejo. Para un proceso univariado AR(1), $y_t = \alpha y_{t-1} + u_t$ esta propiedad significa que

$$1 - \alpha z \neq 0 \quad \text{para } |z| \leq 1$$

o, equivalentemente, $|\alpha| < 1$. Consideremos el caso límite, donde $\alpha = 1$. El proceso resultante $y_t = y_{t-1} + u_t$ se denomina *caminata aleatoria*. Comenzando el proceso en $t = 0$ con algún y_0 fijo, es fácil ver por sustitución sucesiva de los retardos de y_t , que

$$y_t = y_{t-1} + u_t = y_{t-2} + u_{t-1} + u_t = \dots = y_0 + \sum_{i=1}^t u_i.$$

Luego, y_t consiste en la suma de todas las innovaciones de los períodos previos de manera tal que cada innovación tiene un impacto duradero en el proceso. Si u_t es un ruido blanco con varianza σ_u^2 ,

$$E(y_t) = y_0$$

y

$$\text{Var}(y_t) = t\text{Var}(u_t) = t\sigma_u^2,$$

entonces, la varianza de la caminata aleatoria tiende a infinito. Además, se puede ver que $\lim_{t \rightarrow \infty} \text{Corr}(y_t, y_{t+h}) = 1$ para cualquier entero h . Esta última propiedad de una caminata aleatoria significa que y_t e y_s están fuertemente correlacionadas incluso si la distancia entre s y t es grande. Es frecuente que estas propiedades se reflejan en comportamientos de tendencia. Algunos ejemplos se muestran en la Figura 3.3. Esta clase de tendencia es, por supuesto, no una determinística sino una estocástica.

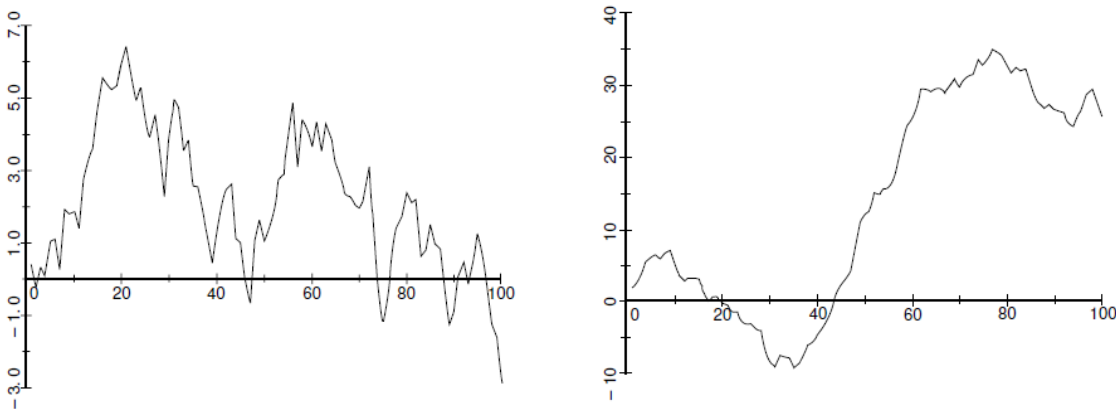


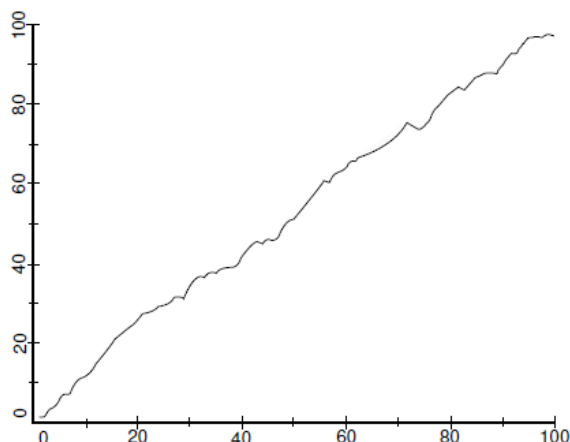
Figura 3.3: Ejemplos de caminata aleatoria.

Si el proceso tiene un término constante distinto de cero ν , $y_t = \nu + y_{t-1} + u_t$ se denomina *caminata aleatoria con drift* y tiene una tendencia lineal determinística en la media. Para ver esta propiedad, supongamos nuevamente que el proceso comienza en $t = 0$ con y_0 fijo. Luego,

$$y_t = y_0 + t\nu + \sum_{i=1}^t u_i$$

y $E(y_t) = y_0 + t\nu$. Un ejemplo de una serie de tiempo generada por una caminata aleatoria con drift se muestra en la Figura 3.4.

Figura 3.4: Ejemplo de caminata aleatoria con drift.



Un comportamiento similar al de una caminata aleatoria también se observa para procesos AR de orden superior tales como

$$y_t = \nu + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + u_t,$$

si $1 - \alpha_1 z - \dots - \alpha_p z^p$ tiene una raíz en $z = 1$. Es decir, si el polinomio característico inverso de y_t tiene una raíz unitaria.

Notar que

$$1 - \alpha_1 z - \dots - \alpha_p z^p = (1 - \lambda_1 z) \dots (1 - \lambda_p z),$$

donde $\lambda_1, \dots, \lambda_p$ son los recíprocos de las raíces de este polinomio. Si el proceso tiene solo una *raíz unitaria* (una raíz igual a 1) y todas las demás raíces están fuera del círculo unitario complejo, su comportamiento es similar al de una caminata aleatoria, es decir, sus varianzas aumentan linealmente, la correlación entre las variables en t y $t + h$ tiende a 1 cuando $t \rightarrow \infty$ y el proceso tiene una tendencia lineal en la media si $\nu \neq 0$. Luego, el proceso es no estacionario.

Por otro lado, en caso de que una de las raíces esté estrictamente dentro del círculo unitario, el proceso se vuelve explosivo, es decir, las varianzas $Var(y_t)$ tienden a infinito exponencialmente en t . Muchos investigadores consideran que dichos procesos son modelos poco realistas para la mayoría de los datos económicos. Aunque los procesos con raíces estrictamente dentro del círculo unitario en ciertos casos son útiles, nos concentraremos en el caso de las raíces unitarias y todas las demás raíces fuera del círculo unitario. Esta situación es de considerable interés práctico.

No se puede aplicar econometría convencional a las series no estacionarias. En particular se puede mostrar que un modelo de regresión podría falsamente indicar una correlación entre dos procesos independientes no estacionarios. Tales regresiones se denominan comúnmente “regresiones espurias”.

Luego, es necesario realizar alguna modificación en la forma de modelar nuestro proceso de manera tal que trabajemos con procesos estables o estacionarios. Para ello, comenzamos dando las siguientes definiciones:

Definición 3.2.1. (*AR(p) integrado*) Un proceso y_t *AR(p)* se dice integrado de orden d , $d \geq 0$, se denota $I(d)$, si y_t tiene exactamente d raíces unitarias y el resto de raíces se encuentra estrictamente fuera del círculo unitario.

Definición 3.2.2. (*VAR(p) integrado*) Un proceso y_t *VAR(p)* se dice integrado de orden d , $x_t \sim I(d)$ si $\Delta^d y_t$ es estable y $\Delta^{d-1} y_t$ no lo es.

En particular, un proceso estable y estacionario se denomina $I(0)$.

Proposición 3.2.3. Si un proceso y_t tiene d raíces unitarias entonces el proceso $\Delta^d y_t$ es estable.

Demostración.

- Caso y_t *AR(p)*: Si $y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + u_t$ es un proceso *AR(p)* con una sola raíz unitaria (y el resto fuera del círculo unitario), es decir, si el proceso es $I(1)$, es posible obtener un proceso estable y posiblemente estacionario a partir de y_t . En efecto, tomando las primeras diferencias $\Delta y_t = y_t - y_{t-1}$ del proceso original.

Para ver esto, consideremos el operador $a(L) \doteq 1 - \alpha_1 L - \dots - \alpha_p L^p = (1 - \lambda_1 L)(1 - \lambda_2 L) \dots (1 - \lambda_p L)$ con $\{\lambda_i\}_i$ las raíces del polinomio característico inverso y L el operador de retardo. Luego,

$$a(L)y_t = u_t.$$

Pero si y_t tiene una única raíz unitaria, esto implica que el polinomio característico inverso $p(z) = 1 - \alpha_1 z - \alpha_2 z^2 - \dots - \alpha_p z^p = (1 - \lambda_1 z)(1 - \lambda_2 z) \dots (1 - \lambda_p z)$ tiene una raíz en $z = 1$ y el resto fuera del círculo unitario. Luego, algún λ_i tiene que ser igual a 1. Supongamos que $\lambda_p = 1$. Entonces, podemos escribir $a(L)$ de la siguiente manera:

$$a(L) = (1 - \lambda_1 L)(1 - \lambda_2 L) \dots (1 - \lambda_{p-1} L)(1 - L).$$

Por lo tanto,

$$a(L)y_t = u_t \Leftrightarrow (1 - \lambda_1 L) \dots (1 - \lambda_{p-1} L)(1 - L)y_t = u_t$$

tal que

$$(1 - \lambda_1 L) \dots (1 - \lambda_{p-1} L)(y_t - y_{t-1}) = (1 - \lambda_1 L) \dots (1 - \lambda_{p-1} L)\Delta y_t = u_t$$

Luego, como $\Delta y_t = (1 - L)y_t$ resulta un proceso $\text{AR}(p)$ estable, dicho proceso es estacionario.

De igual manera, si el proceso es $I(d)$ se puede estabilizar diferenciando d veces. Luego, $\Delta^d y_t = (1 - L)^d y_t$ es estable y, nuevamente, los valores iniciales pueden elegirse de modo que sea estacionario.

- Caso y_t $\text{VAR}(p)$: Consideramos ahora un proceso $\text{VAR}(p)$ integrado de orden d ($I(d)$) de dimensión K definido como en (3.14) y el operador $A(L) \doteq I_K - A_1 L - \dots - A_p L^p$ donde L es el operador de retardo. Luego, podemos escribir

$$A(L)y_t = u_t. \quad (3.15)$$

Multiplicando a izquierda por la adjunta de $A(L)$ que llamamos $A^{adj}(L)$, obtenemos

$$\det(A)y_t = A^{adj}(L)u_t. \quad (3.16)$$

Luego, el proceso $\text{VAR}(p)$ en (3.15) se puede escribir como un proceso con operador AR univariado, es decir, todas las componentes comparten el mismo operador $\det(A(L))$.

El lado derecho en la ecuación anterior, $A^{adj}(L)u_t$, es un proceso de media móvil de orden finito.

Debido a que el proceso es $I(d)$, su polinomio característico inverso $p(z) = \det(I - A_1 z - \dots - A_p z^p)$ tiene exactamente d raíces iguales a 1 y el resto fuera del círculo unitario. Luego, podemos escribir $p(z) = (1 - \lambda_1 z) \dots (1 - \lambda_{p-d} z^{p-d})(1 - z)^d$, donde $\lambda_1, \dots, \lambda_{p-d}$ son las inversas de las raíces de dicho polinomio.

Luego, el operador AR se puede escribir como

$$\det(A(L)) = p(L) = (1 - \lambda_1 L) \dots (1 - \lambda_{p-d} L^{p-d})(1 - L)^d = \alpha(L)(1 - L)^d = \alpha(L)\Delta^d.$$

Debido a que estamos asumiendo $\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_{p-d}}$ fuera del círculo unitario, por Definición 3.1.4, $\alpha(L)$ es un operador invertible. En consecuencia, $\Delta^d y_t = \alpha(L)^{-1} A^{adj}(L)u_t$ lo que implica que cada componente de y_t resulta estable una vez que ésta es diferenciada.

Debido a que estamos considerando procesos que comienzan en un tiempo específico t_0 , debemos analizar este caso cuando multiplicamos por un operador tal como $A(L)^{adj}$ en (3.16). Una posible suposición es que la nueva representación sea válida para todo t tal que el proceso y_t esté definido como en (3.16).

□

Nota. Debido a cancelaciones, puede no ser necesario diferenciar cada componente tantas veces como raíces unitarias posea $\det(A(L))$. Para ilustrar este punto, considere el proceso bivariado $\text{VAR}(1)$

$$\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} L \right) \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} (1 - L)y_{1t} \\ (1 - L)y_{2t} \end{bmatrix} = u_t.$$

Obviamente, cada componente es estacionaria luego de diferenciar una vez, es decir, cada componente es $I(1)$, aunque

$$\det(A) = \det \left(\begin{bmatrix} 1-L & 0 \\ 0 & 1-L \end{bmatrix} \right) = (1-L)^2$$

tiene 2 raíces unitarias.

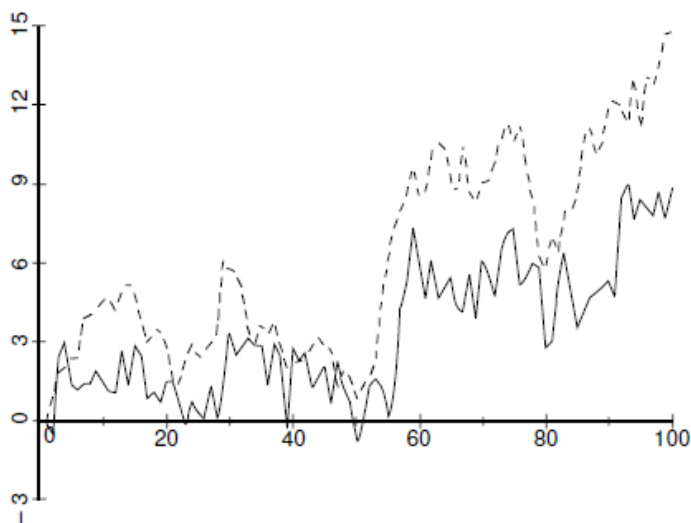
3.3. Procesos cointegrados y modelo de corrección de error

Existen muchas variables económicas de las que se suele sospechar que involucran relaciones de equilibrio. Por ejemplo, los ingresos y gastos de los hogares o los precios de un mismo bien en diferentes mercados.

Supongamos que las variables de interés se recopilan en un vector $y_t = (y_{1t}, \dots, y_{Kt})'$ y que su relación de equilibrio a largo plazo es $\beta' y_t = \beta_1 y_{1t} + \dots + \beta_K y_{Kt} = 0$ donde $\beta = (\beta_1, \dots, \beta_K)$. En cualquier período particular, esta relación puede no ser satisfecha exactamente, pero podemos tener $\beta' y_t = z_t$ donde z_t es un proceso estocástico que representa las desviaciones desde el equilibrio.

Si realmente existe un equilibrio, parece plausible suponer que los procesos y_t se muevan juntos y que z_t sea estable. Por lo tanto, pueden estar impulsados por una *tendencia estocástica común*. Es decir, no se excluye que cada proceso esté integrado, aunque existe una combinación lineal de los procesos que es estacionaria. Los procesos integrados con esta propiedad se llaman *cointegrados*. En la siguiente Figura 3.5, se representan dos series de tiempo cointegradas generadas artificialmente.

Figura 3.5: Ejemplo de Procesos Cointegrados.



Definición 3.3.1. Sea y_t un proceso $I(d)$ tal que $\Delta^d y_t$ es estable y $\Delta^{d-1} y_t$ no lo es. Tal proceso se denomina cointegrado si existe una combinación lineal $\beta' y_t$ con $\beta \neq 0$ que está integrado de orden menor que d .

Por ejemplo, si todos los componentes de y_t son $I(1)$ y βy_t es estacionario ($I(0)$), entonces y_t está cointegrado. El vector β se denomina *vector de cointegración*. Un proceso que consiste en variables cointegradas se denomina *proceso cointegrado*. Estos procesos fueron introducidos por Granger [29] y Engle and Granger [25]. Desde entonces se han hecho populares en el trabajo econométrico teórico y aplicado.

Cabe destacar que nuestra definición de cointegración difiere de la dada por Engle and Granger [25] pues no excluimos componentes de y_t con orden de integración menores que d . En el mencionado trabajo, se requiere que todas las componentes sean $I(d)$.

Aquí, si hay sólo una componente $I(d)$ en y_t y el resto de componentes son estables ($I(0)$), entonces el vector y_t es $I(d)$ de acuerdo a nuestra definición debido a que $\Delta^d y_t$ es estable y $\Delta^{d-1} y_t$ no lo es. En tal caso, una relación $\beta' y_t$ que involucra sólo componentes estacionarias es una relación de cointegración en nuestros términos. Claramente, este aspecto de nuestra definición no está en línea con la idea original de cointegración como una relación especial entre variables integradas con tendencias estocásticas comunes. En lo que sigue, nuestra definición sigue siendo útil porque simplifica la terminología ya que evita distinguir entre variables con diferentes órdenes de integración. Sin embargo, el lector debe tener en cuenta las ideas básicas de cointegración cuando se trata de interpretar relaciones específicas.

Obviamente, un vector de cointegración no es único. Al multiplicarlo por una constante distinta de cero, se obtiene otro vector de cointegración. Además, puede haber varios vectores de cointegración linealmente independientes. Por ejemplo, si hay cuatro variables en un sistema, las dos primeras pueden estar conectadas por una relación de equilibrio a largo plazo y también las dos últimas. Por lo tanto, puede haber un vector de cointegración con ceros en las dos últimas posiciones y uno con ceros en las dos primeras posiciones. Además, puede haber una relación de cointegración que involucre las cuatro variables.

3.3.1. Modelo de corrección de error

Si las relaciones de cointegración están presentes en un sistema de variables, la forma de VAR no es la configuración de modelo más conveniente. En ese caso, es útil considerar parametrizaciones específicas que respalden el análisis de la estructura de cointegración. Los modelos resultantes se conocen como modelos vectoriales de corrección de errores (VECM) o modelos vectoriales de corrección de equilibrio. Estos modelos fueron discutidos en la literatura econométrica antes de que se introdujera el concepto de cointegración y se basan en el supuesto de que los cambios en una variable dependen de las desviaciones de alguna relación de equilibrio.

A continuación, analizaremos esta conexión entre cointegración y VECM's.

A modo de motivación, consideramos el caso particular de un modelo VAR(2) K -dimensional

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + u_t \quad (3.17)$$

con $y_t = (y_{1t}, \dots, y_{Kt})'$ e integrado de orden 1 o $I(1)$. Supongamos que el proceso es inestable con

$$\det(I_K - A_1 z - A_2 z^2) = (1 - \lambda_1 z) \cdots (1 - \lambda_n z) = 0 \quad \text{para } z = 1. \quad (3.18)$$

Debido a que los λ_i son los recíprocos de las raíces del polinomio en (3.18), una o más de ellas deben ser igual a 1. Se supone que todas las demás raíces se encuentran fuera del círculo unitario, es decir, todos los λ_i que no son 1 están dentro del círculo unitario complejo.

Luego, puesto que $\det(I_K - A_1 - A_2) = 0$, la matriz

$$\Pi \doteq -(I_K - A_1 - A_2)$$

es singular. Si A es cualquier matriz, denotaremos su rango con $rk(A)$. Sea r el rango de Π , esto es, $rk(\Pi) = r$. Entonces $r < K$ y Π se puede descomponer como $\Pi = \alpha\beta'$, donde α y β son matrices ($K \times r$) tal que $rk(\alpha) = rk(\beta) = r$ donde rk denota el rango de la matriz. De la discusión en la sección anterior, sabemos que cada variable se vuelve estacionaria al diferenciarse. Luego, si restamos y_{t-1} en ambos lados de (3.17) obtenemos

$$y_t - y_{t-1} = -(I_K - A_1 - A_2)y_{t-1} - A_2 y_{t-1} + A_2 y_{t-2} + u_t.$$

Agrupando los términos con diferencias esto es equivalente a:

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + u_t, \quad \text{donde } \Gamma_1 \doteq -A_2, \quad (3.19)$$

Luego, si $\Pi = \alpha\beta'$ resulta

$$\alpha\beta' y_{t-1} = \Delta y_t - \Gamma_1 \Delta y_{t-1} - u_t.$$

Como Δy_t y Δy_{t-1} no presentan tendencia estocástica debido a nuestra suposición de que todas las variables son a lo sumo $I(1)$, el lado derecho sólo involucra términos estacionarios. Luego, $\alpha\beta' y_{t-1}$ también debe ser estacionario. Ahora, como el $rk(\alpha) = r$, tenemos que $\alpha'\alpha$ es inversible. De esta manera, podemos multiplicar por $(\alpha'\alpha)^{-1}\alpha'$ a izquierda en la última ecuación y obtenemos

$$(\alpha'\alpha)^{-1}\alpha'\alpha\beta' y_{t-1} = (\alpha'\alpha)^{-1}\alpha'\Delta y_t - (\alpha'\alpha)^{-1}\alpha'\Gamma_1 \Delta y_{t-1} - (\alpha'\alpha)^{-1}\alpha' u_t$$

lo cual es equivalente a

$$\beta' y_{t-1} = (\alpha'\alpha)^{-1}\alpha'\Delta y_t - (\alpha'\alpha)^{-1}\alpha'\Gamma_1 \Delta y_{t-1} - (\alpha'\alpha)^{-1}\alpha' u_t.$$

Esto demuestra que $\beta' y_t$ es estacionario y, por lo tanto, cada elemento de $\beta' y_t$ representa una relación de cointegración.

Observación. Hay que tener en cuenta que se podrían tomar las primeras diferencias en todas las variables en (3.17) y ya nos quedarían estacionarias. Sin embargo, de esa manera no podríamos rescatar las relaciones entre las variables que se ven reflejadas en el término de cointegración $\beta' y_t$, las cuales podrían ser de gran importancia. Además, en general, un proceso VAR con variables cointegradas no admite una representación VAR pura en las primeras diferencias.

También es interesante destacar que aquí se ha trabajado bajo el supuesto de que todas las variables son estacionarias después de diferenciar una vez. Sin embargo, pueden existir variables con órdenes de integración superior y, en ese caso, $\beta' y_t$ puede no ser estacionario incluso si $rk(\Pi) = r < K$. Las componentes de y_t aún pueden ser cointegradas de un orden superior si existen combinaciones lineales que tienen un orden de integración reducido.

A continuación, nos interesará el caso particular donde las componentes del proceso y_t son $I(1)$ ó $I(0)$.

Definición 3.3.2. *El proceso K -dimensional VAR(p)*

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \quad (3.20)$$

se denomina cointegrado de rango r si

$$\Pi \doteq -(I_K - A_1 - \cdots - A_p)$$

tiene rango r y, por lo tanto, Π se puede escribir como un producto $\alpha\beta'$ con α y β de dimensión $(K \times r)$ y de rango r .

La matriz β se denomina matriz de cointegración y α se denomina matriz de carga. Por último, las matrices Γ_j se suelen denominar parámetros de corto plazo y Πy_{t-1} se suele llamar la parte de largo plazo.

Reescribiendo (3.20) como en (3.19), obtenemos una representación de Modelo de Corrección de Error ó VECM para y_t de la siguiente manera:

$$\begin{aligned} \Delta y_t &= \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t \\ &= \alpha\beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t, \end{aligned} \quad (3.21)$$

donde

$$\Gamma_i \doteq -(A_{i+1} + \cdots + A_p), \quad \text{para } i = 1, \dots, p-1.$$

Si esta representación de un proceso cointegrado es dada, es fácil recuperar la forma VAR correspondiente (3.20) notando que

$$\begin{aligned} A_1 &= \Pi + I_K + \Gamma_1, \\ A_i &= \Gamma_i - \Gamma_{i-1} \quad \text{para } i = 2, \dots, p-1, \\ A_p &= -\Gamma_{p-1}. \end{aligned}$$

De la definición previa podemos analizar tres casos de importancia:

- $rk(\Pi) = 0$. Esto implica que $\Pi = 0$ e y_t no está cointegrado. Luego, el VECM se reduce a

$$\Delta y_t = \Gamma_1 \Delta y_{t-1} + \dots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t,$$

tal que Δy_{t-1} sigue un modelo VAR($p-1$). Es decir, Δy_t tiene una representación VAR estable. En otras palabras, existe una representación VAR estable para las primeras diferencias de las variables en lugar de los niveles de las variables.

- $rk(\Pi) = K$. Esto implica que $\det(\Pi) = \det(I_K - A_1 - \dots - A_p) \neq 0$ e y_t no contiene raíces unitarias; es decir, y_t es $I(0)$. El VECM no es informativo y uno estudia directamente y_t .
- $0 < rk(\Pi) = r < K$. En este caso, uno puede escribir a Π como $\Pi = \alpha\beta'$, donde α y β son matrices $K \times r$ con $rk(\alpha) = rk(\beta) = r$. Luego, el VECM resulta como en (3.21). Esto significa que y_t está cointegrado con r vectores de cointegración linealmente independientes, $w_t = \beta' y_t$ y tiene $K - r$ raíces unitarias que dan $K - r$ tendencias estocásticas comunes de y_t .

Observación. También hay otros casos en los que no existe cointegración en el sentido original, aunque el modelo (3.21) tenga un rango de cointegración estrictamente entre 0 y K . Supongamos, por ejemplo, que todas las variables excepto una son $I(0)$; luego, el rango de cointegración es $K - 1$, aunque la variable $I(1)$ no está cointegrada con las otras variables. De forma similar, podría haber $K - r$ variables $I(1)$ no relacionadas y r componentes $I(0)$. Generalmente, para cada variable $I(0)$ en el sistema puede haber una columna en la matriz β con un 1 en una posición y ceros en el resto de posiciones.

Estos casos no representan una relación de cointegración en el sentido original del término. Sin embargo, es conveniente incluir estos casos en el presente marco porque se pueden acomodar fácilmente en lo que respecta a la estimación y la inferencia.

Por otro lado, es importante observar que la descomposición de la matriz Π ($K \times K$) como el producto de dos matrices ($K \times r$), $\Pi = \alpha\beta'$, no es única. De hecho, para cada matriz no singular Q ($r \times r$), podemos definir $\alpha^* = \alpha Q'$ y $\beta^* = \beta Q^{-1}$ y obtener $\Pi = \alpha^* \beta^{*'}$. Esta no unicidad de la descomposición de Π muestra nuevamente que las relaciones de cointegración no son únicas. Sin embargo, es posible imponer

restricciones sobre β y/o α para obtener relaciones únicas. Dichas restricciones pueden estar implícitas por consideraciones del tema en particular o pueden imponerse por conveniencia, utilizando las propiedades algebraicas de las matrices asociadas.

Usando el hecho de que $rk(\beta) = r$, la matriz β debe tener r filas linealmente independientes. Por lo tanto, mediante una reorganización adecuada de las variables siempre se puede garantizar que las primeras r filas de β sean linealmente independientes. Por lo tanto, la submatriz superior ($r \times r$) que consta de las primeras r filas de β es no singular. Luego, podemos considerar Q de dimensión ($r \times r$) como la inversa de esta submatriz y obtener la siguiente matriz de cointegración

$$\beta^* = \beta Q^{-1} = \begin{pmatrix} I_r \\ \beta_{(K-r)} \end{pmatrix}, \quad (3.22)$$

donde $\beta_{(K-r)}$ es $((K-r) \times r)$. En general, usaremos esta normalización porque es muy conveniente dar condiciones para definir una matriz de cointegración única. Esto no implica una pérdida de generalidad, excepto que se supone que las variables están organizadas de la manera correcta para que la normalización sea factible.

El resultado importante es que la normalización de la matriz de cointegración dada en la fórmula (3.22) siempre es posible si las variables están organizadas de una manera adecuada. Encontrar el orden correcto es fácil si se conocen las propiedades de cointegración de todos los subsistemas (subconjunto de variables), incluidos los subsistemas univariados. En otras palabras, también necesitamos saber el orden de integración de las variables individuales en el sistema.

En la práctica, el orden de integración y el rango de cointegración de un sistema dado y sus subsistemas no se conocen. Sin embargo, existen procedimientos estadísticos para determinar el rango de cointegración que pueden ayudar a superar este problema práctico. Veremos esto en las siguientes secciones.

3.3.2. Modelo de corrección de error con término determinista

Es frecuente que los términos deterministas estén presentes en procesos cointegrados y VECM y que produzcan diferentes impactos en los mismos así como en las distribuciones límites de los test de cointegración.

De hecho, a partir de la discusión de la caminata aleatoria con drift, debe quedar claro que los términos deterministas en un proceso VAR con raíces unitarias pueden tener un impacto diferente que en un VAR estable. Por ejemplo, un término de intersección en una caminata aleatoria genera una tendencia lineal en la media del proceso, mientras que un término de intersección en un proceso de AR estable simplemente implica un valor medio constante.

Para explorar las implicaciones del término determinista, se asume el siguiente modelo:

$$y_t = \mu_t + x_t, \quad (3.23)$$

donde x_t es un proceso VAR(p) de media cero con variables posiblemente cointegradas y μ_t representa el término determinista. Por ejemplo, el término determinista puede ser simplemente una constante, $\mu_t = \mu_0$, o puede ser un término de tendencia lineal, $\mu_t = \mu_0 + \mu_1 t$, donde μ_0 y μ_1 son vectores de parámetros K -dimensionales fijos. Otros posibles términos deterministas que pueden incluirse son variables ficticias estacionales entre otras, para dar cuenta de eventos especiales. La ventaja de configurar el proceso en la forma (3.23) agregando la parte determinista a la parte estocástica de media cero es que la media de las variables y_t está claramente especificada por el término determinista y no necesita derivarse de cantidades que involucran además los parámetros de la parte estocástica. La desventaja es que la parte estocástica x_t no es directamente observable en general. Por lo tanto, para fines de estimación, por ejemplo, tenemos que reescribir el proceso en términos de los y_t observables. Lo haremos a continuación para algunos casos de interés específico.

Supongamos que el proceso x_t se puede representar en forma de VECM como en (3.21). Considerando ahora el caso de un término determinista constante, $\mu_t = \mu_0$, tenemos $x_t = y_t - \mu_0$ de modo que $\Delta y_t = \Delta x_t$ y de (3.21) para x_t obtenemos:

$$\begin{aligned} \Delta y_t &= \Pi(y_{t-1} - \mu_0) + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t \\ &= \alpha \beta^{0'} \begin{bmatrix} y_{t-1} \\ 1 \end{bmatrix} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t \\ &= \Pi^0 y_{t-1}^0 + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t, \end{aligned} \quad (3.24)$$

donde $\beta^{0'} \doteq [\beta', \tau']$ con $\tau' \doteq -\beta' \mu_0$ un vector ($r \times 1$), $y_{t-1}^0 = \begin{bmatrix} y_{t-1} \\ 1 \end{bmatrix}$ y $\Pi_0 \doteq [\Pi : \nu_0]$ es ($K \times (K + 1)$) con $\nu_0 \doteq -\Pi \mu_0 = \alpha \tau'$. Por lo tanto, si sólo hay una media constante, se puede absorber en las relaciones de cointegración. En otras palabras, la media constante se convierte en un término de intercepto en las relaciones de cointegración. Por supuesto, el modelo también se puede escribir de la forma:

$$\begin{aligned} \Delta y_t &= \nu_0 + \alpha \beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t \\ &= \nu_0 + \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t, \end{aligned} \quad (3.25)$$

donde ν_0 se denomina término de intercepto general. Aquí ν_0 no puede ser un vector arbitrario ($K \times 1$) sino que tiene que satisfacer las restricciones indicadas ($\nu_0 = \alpha \tau'$) para asegurar que el término de intercepto en este modelo no genere una tendencia lineal en la media de las variables y_t . Al especificar el término determinista en forma aditiva como en (3.23), las propiedades de la media de y_t son fáciles de ver.

Un proceso con una tendencia lineal en la media, $\mu_t = \mu_0 + \mu_1 t$, es otro caso de importancia práctica. Usando $x_t = y_t - \mu_0 - \mu_1 t$, $\Delta x_t = \Delta y_t - \mu_1$ y (3.21) para x_t , se obtiene:

$$\Delta y_t - \mu_1 = \alpha \beta' (y_{t-1} - \mu_0 - \mu_1(t-1)) + \Gamma_1 (\Delta y_{t-1} - \mu_1) + \dots \quad (3.26)$$

$$+ \Gamma_{p-1} (\Delta y_{t-p+1} - \mu_1) + u_t \quad (3.27)$$

o recopilar términos deterministas

$$\begin{aligned} \Delta y_t &= \nu + \alpha [\beta' : \eta'] \begin{bmatrix} y_{t-1} \\ t-1 \end{bmatrix} + \Gamma_1 \Delta y_{t-1} + \dots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t, \\ &= \nu + \Pi^+ y_{t-1}^+ + \Gamma_1 \Delta y_{t-1} + \dots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t, \end{aligned} \quad (3.28)$$

donde $\nu \doteq -\Pi \mu_0 + (I_K - \Gamma_1 - \dots - \Gamma_{p-1}) \mu_1$, $\eta' \doteq -\beta' \mu_1$, $\Pi^+ \doteq \alpha [\beta' : \eta']$ es una matriz $(K \times (K+1))$ y $y_t^+ \doteq \begin{bmatrix} y_t \\ 1 \end{bmatrix}$.

Ahora, el término de intercepto general ν no tiene restricciones y puede tomar cualquier valor de \mathbb{R}^K , dependiendo, por supuesto, de μ_0 , μ_1 y los demás parámetros. Por el contrario, el término tendencia se puede absorber en las relaciones de cointegración. Escribiendo el modelo con un término de tendencia lineal sin restricciones en la forma

$$\Delta y_t = \nu_0 + \nu_1 t + \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \dots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t,$$

el modelo es realmente, en principio, capaz de generar tendencias cuadráticas en las medias de las variables.

También es posible que el parámetro μ_1 representando la pendiente de la tendencia sea ortogonal a la matriz de cointegración de modo que $\beta' \mu_1 = 0$ y, por tanto, $\nu = 0$ y el término de tendencia desaparezca de las relaciones de cointegración. Esta situación también puede ocurrir si $\mu_1 \neq 0$ y las variables tienen tendencias lineales en sus medias. Entonces, las tendencias lineales se generarán mediante el término de intersección ν . Esto representa una situación donde aparece una tendencia lineal en las variables pero no en las relaciones de cointegración. Sin embargo, observe que en esta situación el rango de cointegración debe ser menor que K . Si el proceso tiene un rango de cointegración K , es estable y, por lo tanto, no puede generar una tendencia lineal cuando sólo se incluye un intercepto en el modelo. Formalmente, una “matriz de cointegración” β de rango K es no singular, de modo que $\beta' \mu_1$ no puede ser cero si μ_1 es distinto de cero.

Este análisis se amplía más adelante en la Sección 3.3.4.

3.3.3. Teorema de Representación de Granger

Finalizamos esta sección dando condiciones necesarias en los coeficientes del proceso modelado a través de un VAR para que dicho proceso sea $I(1)$. Para este propósito, Johansen [35] considera una

reparametrización del modelo VAR como un Modelo de Corrección de Error, establece las condiciones para que el proceso autorregresivo sea $I(1)$ y deriva su representación de media móvil.

Este resultado a menudo se denomina *Teorema de Representación de Granger* y lo presentamos a continuación:

Teorema 3.3.3 (Teorema de Representación de Granger). *Sea y_t un proceso K -dimensional $I(1)$ y cointegrado de rango r tal que*

$$\Delta y_t = \alpha \beta' y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \phi d_t + u_t, \quad (3.29)$$

donde $y_t = 0$ para todo $t \leq 0$, u_t es ruido blanco para $t = 1, 2, \dots$ con $u_t = 0$ para $t \leq 0$ y d_t es un término determinista que puede contener una constante, un término lineal, variables estacionales, entre otros regresores fijos y no estocásticos. Además, definimos

$$C(z) \doteq (1 - z)I_K - \alpha \beta' z - \sum_{i=1}^{p-1} \Gamma_i (1 - z) z^i \quad (3.30)$$

y asumimos ciertas las siguientes condiciones para los parámetros:

- (a) $\det(C(z)) = 0 \Rightarrow |z| > 1$ ó $z = 1$.
- (b) El número de raíces unitarias, $z = 1$, es exactamente $K - r$.
- (c) α y β son matrices $K \times r$ con $rk(\alpha) = rk(\beta) = r$.

Entonces, y_t tiene la representación

$$y_t = \Psi \sum_{i=1}^t (u_i + \phi d_i) + \Psi^*(L)(u_t + \phi d_t) + y_0^* \quad (3.31)$$

donde

$$\Psi = \beta_{\perp} \left[\alpha'_{\perp} \left(I_K - \sum_{i=1}^{p-1} \Gamma_i \right) \beta_{\perp} \right]^{-1} \alpha'_{\perp},$$

con $\Psi^*(L)u_t = \sum_{j=0}^{\infty} \Psi_j^* u_{t-j}$ un proceso $I(0)$ e y_0^* conteniendo valores iniciales. Más aún, y_0^* satisface $\beta' y_0^* = 0$.

Demostración.

Sea y_t un proceso VAR(p) K -dimensional, tal que

1. cada componente es $I(1)$ o $I(0)$.

2. Satisface (3.29) donde $y_t = 0$ para $t < 0$, u_t es un ruido blanco para $t \in \mathbb{N}$ y $u_t = 0$ para $t < 0$.

3. El polinomio

$$C(z) := (1 - z)I_K - \alpha\beta'z - \sum_{j=1}^{p-1} \Gamma_j(1 - z)z^j,$$

cumple que:

- a) Si $|C(z)| = 0$ entonces $|z| > 1$ o $z = 1$.
- b) El número de raíces unitarias de $|C(z)|$ es exactamente $K - r$.
- c) α y β son matrices $K \times r$ de rango r .

En primer lugar, notamos que

$$\Delta y_t = \alpha\beta'y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + u_t + \phi d_t$$

se puede escribir como

$$\left((I_k - L) - \sum_{j=1}^{p-1} \Gamma_j (I_k - L)L^j - \alpha\beta'L \right) y_t = u_t + \phi d_t, \quad (3.32)$$

donde $Lx_t = x_{t-1}$ es el operador de retardo. Así, si $C(z)$ es como en la definición, entonces (3.32) se puede escribir como

$$C(L)y_t = u_t + \phi d_t. \quad (3.33)$$

Paso 1. Dado que $|C(z)|$ tiene exactamente $K - r$ raíces unitarias, se busca escribir a $C(z)$ como producto de matrices

$$C(z) = A(z)P(z)$$

donde $|P(z)| = \lambda(1 - z)^{K-r}$, para cierto $\lambda \in \mathbb{R}$, $\lambda \neq 0$. De esta manera $|A(z)|$ tiene todas sus raíces fuera del círculo unitario, y por lo tanto $A(L)$ es un operador invertible.

Para esto, observamos que el rango de β es r , y por lo tanto β' tiene una *inversa* $\bar{\beta}$ a derecha:

$$\beta' \cdot \underbrace{(\beta(\beta'\beta)^{-1})}_{\bar{\beta}} = I_r.$$

Notar que $\bar{\beta}$ es $K \times r$. Además $\bar{\beta}'\beta = I_r$.

Consideramos ahora $K - r$ vectores que generen el espacio ortogonal al generado por las columnas de β , y llamamos β_\perp a la matriz $k \times (k - r)$ formada por estos vectores. Sean Q^{-1} y Q las matrices inversibles $K \times K$ dadas por:

$$Q^{-1} = [\bar{\beta} : \beta_\perp], \quad Q = \begin{bmatrix} \beta' \\ (\bar{\beta}_\perp)' \end{bmatrix}$$

donde $\overline{\beta}_\perp$ es la inversa a derecha de β_\perp . Esto es:

$$\beta'_\perp \cdot \overline{\beta}_\perp = I_{K-r}.$$

Notemos que $\gamma \in \mathbb{R}^K$ es ortogonal a los vectores de β , ($\gamma' \beta = 0$), si y sólo si γ es ortogonal a los vectores de $\overline{\beta}$, pues

$$\gamma' \cdot \overline{\beta} = \gamma' \cdot (\beta (\beta' \beta)^{-1}) = (\gamma' \beta) \cdot (\beta' \beta)^{-1},$$

y $\beta' \beta$ es inversible. Luego:

$$Q \cdot Q^{-1} = \begin{pmatrix} \beta' \\ (\overline{\beta}_\perp)' \end{pmatrix} [\overline{\beta} : \beta_\perp] = \begin{pmatrix} \beta' \overline{\beta} & \beta' \beta_\perp \\ (\overline{\beta}_\perp)' \overline{\beta} & (\overline{\beta}_\perp)' \beta_\perp \end{pmatrix} = \begin{pmatrix} I_r & 0 \\ 0 & I_{k-r} \end{pmatrix}$$

Ahora escribimos:

$$C(z) = (1-z) \underbrace{\left(I_k - \sum_{j=1}^{p-1} \Gamma_j z^j \right)}_{\Gamma(z)} - \alpha \beta' z.$$

$$\begin{aligned} C(z) &= C(z) Q^{-1} Q \\ &= ((1-z)\Gamma(z) - \alpha \beta' z) [\overline{\beta} : \beta_\perp] Q \\ &= [(1-z)\Gamma(z)\overline{\beta} - \alpha \beta' \overline{\beta} z : (1-z)\Gamma(z)\beta_\perp - \alpha \beta' \beta_\perp z] Q \\ &= [(1-z)\Gamma(z)\overline{\beta} - \alpha z : (1-z)\Gamma(z)\beta_\perp] Q \\ &= [(1-z)\Gamma(z)\overline{\beta} - \alpha z : \Gamma(z)\beta_\perp] \begin{bmatrix} I_r & \\ & (1-z)I_{k-r} \end{bmatrix} \begin{bmatrix} \beta' \\ (\overline{\beta}_\perp)' \end{bmatrix} \\ &= \underbrace{[(1-z)\Gamma(z)\overline{\beta} - \alpha z : \Gamma(z)\beta_\perp]}_{A(z)} P(z), \end{aligned}$$

donde

$$P(z) = \begin{pmatrix} I_r & \\ & (1-z)I_{K-r} \end{pmatrix} \begin{pmatrix} \beta' \\ (\overline{\beta}_\perp)' \end{pmatrix} = \begin{pmatrix} \beta' \\ (1-z)(\overline{\beta}_\perp)' \end{pmatrix}$$

es una matriz $K \times K$ y cuyo determinante $\det(P(z)) = (1-z)^{K-r} |Q|$ tiene exactamente $K-r$ raíces, todas iguales a 1.

Finalmente, escribimos

$$C(z) = Q^{-1} Q \underbrace{[(1-z)\Gamma(z)\overline{\beta} - \alpha z : \Gamma(z)\beta_\perp]}_{B_*(z)} P(z) \quad (3.34)$$

donde $|B_*(z)|$ tiene todas sus raíces fuera del círculo unitario. Por lo tanto $B_*(L)$ resulta un operador inversible.

Paso 2. Como siguiente paso, escribimos

$$C(z) = \underbrace{Q^{-1}B_*(z)Q}_{B(z)} Q^{-1}P(z). \quad (3.35)$$

Definimos

$$z_t = Q^{-1}P(L)y_t = [\bar{\beta} : \beta_{\perp}] \begin{bmatrix} \beta' \\ (I-L)(\bar{\beta}_{\perp})' \end{bmatrix} y_t = \bar{\beta}\beta'y_t + \beta_{\perp}(\bar{\beta}_{\perp})'\Delta y_t. \quad (3.36)$$

Entonces

$$\beta'z_t = \beta'y_t.$$

Las raíces del determinante de $B(z)$ están fuera del círculo unitario, por lo cual el operador $B(L)$ es inversible (ver Observación en página 76). Además:

$$B(0) = Q^{-1}B_*(0)Q = Q^{-1}(QQ^{-1})Q = I_K.$$

También

$$B(L)z_t = Q^{-1}B_*(L)QQ^{-1}P(L)y_t = C(L)y_t = u_t + \phi d_t. \quad (3.37)$$

Por lo tanto $B(z) = B_0 - \sum_{i=1}^p B_i z^i$ es un polinomio matricial de orden p , con término independiente $B_0 = B(0)$ igual a I_k . Esto ocurre pues B_* tiene orden p ya que en uno de sus términos aparece $(1-z)\Gamma(z)$, con Γ de orden $p-1$. Luego

$$B(L)z_t = B_0 z_t - \sum_{i=1}^p B_i L^i z_t = z_t - \sum_{i=1}^p B_i z_{t-i} = u_t + \phi d_t,$$

y por lo tanto

$$z_t = \sum_{i=1}^p B_i z_{t-i} + u_t + \phi d_t. \quad (3.38)$$

Como $B(L)$ es un operador invertible, llamamos $\Theta(L) = \sum_{j=0}^{\infty} \Theta_j L^j$ a su inversa. De (3.37) despejamos z_t :

$$z_t = (B(L))^{-1}(u_t + \phi d_t) = \Theta(L)(u_t + \phi d_t) = \sum_{j=0}^{\infty} \Theta_j (u_{t-j} + \phi d_{t-j}). \quad (3.39)$$

Paso 3. Volviendo a la expresión de y_t tenemos que

$$y_t = Q^{-1}Qy_t = \bar{\beta}\beta'y_t + \beta_{\perp}(\bar{\beta}_{\perp})'y_t.$$

$$\Delta y_t = \bar{\beta}\beta'\Delta y_t + \beta_{\perp}(\bar{\beta}_{\perp})'\Delta y_t. \quad (3.40)$$

De la definición de z_t en la ecuación (3.36), y el hecho que $\beta' z_t = \beta' y_t$ tenemos que

$$\begin{aligned} z_t &= \bar{\beta}\beta' y_t + \beta_{\perp}(\bar{\beta}_{\perp})' \Delta y_t && \text{definición de } z_t \\ &= \bar{\beta}\beta' y_t + (\Delta y_t - \bar{\beta}\beta' \Delta y_t) && \text{por (3.40)} \\ &= \bar{\beta}\beta' y_t + \Delta y_t - \bar{\beta}\beta' y_t + \bar{\beta}\beta' y_{t-1} \\ &\quad \Delta y_t + \bar{\beta}\beta' z_{t-1} \end{aligned}$$

es decir que

$$\Delta y_t = z_t - \bar{\beta}\beta' z_{t-1}$$

Entonces:

$$\begin{aligned} \Delta y_t &= \bar{\beta}\beta' \Delta y_t + \beta_{\perp}(\bar{\beta}_{\perp})' \Delta y_t \\ &= \bar{\beta}\beta' \Delta y_t + \beta_{\perp}(\bar{\beta}_{\perp})'(z_t - \bar{\beta}\beta' z_{t-1}) \\ &= \bar{\beta}\beta' \Delta y_t + \beta_{\perp}(\bar{\beta}_{\perp})' z_t \end{aligned}$$

$$\boxed{\Delta y_t = \beta_{\perp}(\bar{\beta}_{\perp})' z_t + \bar{\beta}\beta' \Delta y_t = \omega_t}$$

Con ω_t un proceso estacionario por ser y_t un proceso $I(1)$.

Paso 4. Ahora podemos escribir:

$$y_t = y_{t-1} + \omega_t = y_0 + (\omega_t + \omega_{t-1} + \dots + \omega_1) \quad (3.41)$$

donde

$$\omega_t = \underbrace{\beta_{\perp}(\bar{\beta}_{\perp})' z_t}_{\eta} + \underbrace{\bar{\beta}\beta' \Delta y_t}_{\mu}$$

Así,

$$\begin{aligned} \omega_t &= \eta z_t + \mu \Delta y_t = \eta z_t + \mu(y_t - y_{t-1}) \\ \omega_{t-1} &= \eta z_{t-1} + \mu(y_{t-1} - y_{t-2}) \\ \omega_{t-2} &= \eta z_{t-2} + \mu(y_{t-2} - y_{t-3}) \\ &\vdots = \vdots \\ \omega_2 &= \eta z_2 + \mu(y_2 - y_1) \\ \omega_1 &= \eta z_1 + \mu(y_1 - y_0) \end{aligned}$$

Recordemos que $z_t = B(L)^{-1}(u_t + \phi d_t) = \sum_{j=0}^{\infty} \Theta_j(u_{t-j} + \phi d_{t-j})$, y que $u_t = d_t = 0$ para $t < 0$.

Denotemos por $\tilde{u}_t = u_t + \phi d_t$. Luego tenemos que

$$\begin{aligned} z_t &= \Theta_0 \tilde{u}_t + \Theta_1 \tilde{u}_{t-1} + \cdots + \Theta_{t-2} \tilde{u}_2 + \Theta_{t-1} \tilde{u}_1 + \Theta_t \tilde{u}_0 \\ z_{t-1} &= \Theta_0 \tilde{u}_{t-1} + \Theta_1 \tilde{u}_{t-2} + \cdots + \Theta_{t-2} \tilde{u}_1 + \Theta_{t-1} \tilde{u}_0 \\ &\vdots \\ z_2 &= \Theta_0 \tilde{u}_2 + \Theta_1 \tilde{u}_1 + \Theta_2 \tilde{u}_0 \\ z_1 &= \Theta_0 \tilde{u}_1 + \Theta_1 \tilde{u}_0. \end{aligned}$$

Por lo tanto, agrupando los términos con $\tilde{u}_0, \tilde{u}_1, \tilde{u}_2, \dots, \tilde{u}_t$:

$$\begin{aligned} \omega_t + \omega_{t-1} + \cdots + \omega_1 &= \eta(z_t + \cdots + z_1) + \mu(y_t - y_0) \\ &= \eta((\Theta_1 + \cdots + \Theta_t) \tilde{u}_0 + (\Theta_0 + \cdots + \Theta_{t-1}) \tilde{u}_1 + \cdots + (\Theta_0 + \Theta_1) \tilde{u}_{t-1} + \Theta_0 \tilde{u}_t) \\ &\quad + \mu(y_t - y_0). \end{aligned} \tag{3.42}$$

Ahora escribimos:

$$\sum_{j=0}^k \Theta_j = \underbrace{\sum_{j=0}^{\infty} \Theta_j}_{\Theta(1)} + \underbrace{\left(-\sum_{j=k+1}^{\infty} \Theta_j\right)}_{\Theta_k^*}.$$

Cada Θ_k^* está bien definido pues son los coeficientes de $\Theta'(1)$.

Así, (3.42) se puede reescribir:

$$\begin{aligned} \omega_1 + \cdots + \omega_t &= \eta((\Theta(1) + \Theta_0^*) \tilde{u}_t + (\Theta(1) + \Theta_1^*) \tilde{u}_{t-1} + \cdots + (\Theta(1) + \Theta_t^* - \Theta_0) \tilde{u}_0) \\ &\quad + \mu(y_t - y_0)) \\ &= \eta(\Theta(1)(\tilde{u}_1 + \tilde{u}_2 + \cdots + \tilde{u}_t)) + \eta\left(\sum_{j=0}^{\infty} \Theta_j^* \tilde{u}_{t-j}\right) \\ &\quad + \mu(y_t - y_0) + \eta(\Theta(1) - \Theta_0) \tilde{u}_0. \end{aligned} \tag{3.43}$$

Notemos que como $u_t = d_t = 0$ para $t < 0$, entonces $\sum_{j=0}^{\infty} \Theta_j^* \tilde{u}_{t-j} = \sum_{j=0}^t \Theta_j^* \tilde{u}_{t-j}$, pero para que no quede la sumatoria dependiente de t podemos escribirlo como

$$\sum_{j=0}^{\infty} \Theta_j^* \tilde{u}_{t-j} = \sum_{j=0}^{\infty} \Theta_j^* L^j \tilde{u}_t = \Theta^*(L) \tilde{u}_t.$$

Si no se tuviera la hipótesis $\tilde{u}_t = 0$ para $t \leq t_0$ (algún t_0), debería haber alguna condición sobre Θ_j^* para asegurar convergencia.

Ahora queremos estudiar el término $\mu(y_t - y_0)$. Notemos que

$$\mu y_t = \bar{\beta} \beta' y_t = \bar{\beta} \beta' z_t = \bar{\beta} \beta' \Theta(L) \tilde{u}_t = \mu \Theta(L) \tilde{u}_t.$$

Volviendo a (3.43), tenemos que

$$\begin{aligned}
\omega_1 + \cdots + \omega_t &= \eta \Theta(1) \sum_{j=1}^t \tilde{u}_j + \eta \Theta^*(L) \tilde{u}_t + \mu \Theta(L) \tilde{u}_t - \mu y_0 + \eta(\Theta(1) - \Theta_0) \tilde{u}_0 \\
&= \underbrace{(\beta_{\perp} \overline{\beta_{\perp}}')}_{\Psi} \Theta(1) \sum_{j=1}^t \tilde{u}_j + \underbrace{(\beta_{\perp} \overline{\beta_{\perp}}' \Theta^*(L) + \overline{\beta} \beta' \Theta(L))}_{\Psi^*(L)} \tilde{u}_t \\
&\quad - \overline{\beta} \beta' y_0 + \beta_{\perp} \overline{\beta_{\perp}}' (\Theta(1) - \Theta_0) \tilde{u}_0
\end{aligned} \tag{3.44}$$

Paso 5. Analicemos Ψ . Como $B(z)\Theta(z) = Id$, entonces $B(1)\Theta(1) = Id$. Por lo tanto

$$\Theta(1) = B(1)^{-1}.$$

De (3.34) tenemos que

$$B^*(1) = Q[-\alpha : \Gamma(1)\beta_{\perp}]$$

y de (3.35):

$$B(1) = Q^{-1}Q[-\alpha : \Gamma(1)\beta_{\perp}]Q = [-\alpha : \Gamma(1)\beta_{\perp}]Q.$$

Luego

$$\begin{aligned}
\Psi &= \beta_{\perp} \overline{\beta_{\perp}}' Q^{-1}[-\alpha : \Gamma(1)\beta_{\perp}]^{-1} \\
&= \beta_{\perp} \overline{\beta_{\perp}}' [\beta : \beta_{\perp}] [-\alpha : \Gamma(1)\beta_{\perp}]^{-1} \\
&= \beta_{\perp} [0 : I_{k-r}] [-\alpha : \Gamma(1)\beta_{\perp}]^{-1}
\end{aligned} \tag{3.45}$$

La última igualdad sale del hecho que $(\beta'_{\perp} \beta_{\perp})(\beta'_{\perp} \beta_{\perp})^{-1} = I_{(k-r) \times (k-r)}$, entonces

$$\overline{\beta_{\perp}} = \beta_{\perp} (\beta'_{\perp} \beta_{\perp})^{-1},$$

entonces

$$\overline{\beta_{\perp}}' \beta = ((\beta'_{\perp} \beta_{\perp})^{-1})' \beta'_{\perp} \beta = ((\beta'_{\perp} \beta_{\perp})^{-1})' \cdot 0 = 0.$$

Calculamos ahora $[-\alpha : \Gamma(1)\beta_{\perp}]^{-1}$. La matriz $[-\alpha : \Gamma(1)\beta_{\perp}]$ es $k \times k$, por lo cual su inversa U también lo es ¹. Como α y α_{\perp} generan \mathbb{R}^k , entonces podemos decir que

$$U = \begin{pmatrix} A\alpha' + B\alpha'_{\perp} \\ C\alpha' + D\alpha'_{\perp} \end{pmatrix} \tag{3.46}$$

¹Acá notar que $-\alpha$ es una matriz $K \times r$

donde A , B , C y D son matrices de orden $r \times r$, $r \times (k-r)$, $(k-r) \times r$ y $(k-r) \times (k-r)$ respectivamente. Luego debe cumplirse que

$$U \cdot [-\alpha : \Gamma(1)\beta_{\perp}] = \begin{pmatrix} A\alpha' + B\alpha'_{\perp} \\ C\alpha' + D\alpha'_{\perp} \end{pmatrix} [-\alpha : \Gamma(1)\beta_{\perp}] = \begin{pmatrix} I_r & 0_{r \times (k-r)} \\ 0_{(k-r) \times r} & I_{k-r} \end{pmatrix}. \quad (3.47)$$

Por lo tanto:

$$(A\alpha' + B\alpha'_{\perp}) \cdot (-\alpha) = I_r \quad (3.48)$$

$$(A\alpha' + B\alpha'_{\perp}) \cdot (\Gamma(1)\beta_{\perp}) = 0_{r \times (k-r)} \quad (3.49)$$

$$(C\alpha' + D\alpha'_{\perp}) \cdot (-\alpha) = 0_{(k-r) \times r} \quad (3.50)$$

$$(C\alpha' + D\alpha'_{\perp}) \cdot (\Gamma(1)\beta_{\perp}) = I_{k-r} \quad (3.51)$$

De la ecuación (3.50) tenemos que

$$C\alpha'(-\alpha) + D\alpha'_{\perp}(-\alpha) = 0.$$

Por un lado, $\alpha'_{\perp}\alpha = 0$, luego debe ser $C\alpha'(-\alpha) = 0$. Como el rango de α es r , entonces $\alpha'\alpha$ es una matriz inversible $r \times r$, y por lo tanto

$$C = 0. \quad (3.52)$$

Utilizamos (3.52) en (3.51), y vemos entonces que

$$D\alpha'_{\perp}\Gamma(1)\beta_{\perp} = I_{k-r}, \quad (3.53)$$

de donde se deduce directamente que la matriz cuadrada D es inversible y

$$D = (\alpha'_{\perp}\Gamma(1)\beta_{\perp})^{-1}. \quad (3.54)$$

De (3.48) se desprende que $A\alpha'(-\alpha) = I_r$, por lo cual

$$A = -(\alpha'\alpha)^{-1}. \quad (3.55)$$

En (3.49) resulta entonces

$$\begin{aligned} -(\alpha'\alpha)^{-1}\alpha' + B\alpha'_{\perp})\Gamma(1)\beta_{\perp} &= 0 \\ B\alpha'_{\perp}\Gamma(1)\beta_{\perp} &= (\alpha'\alpha)^{-1}\alpha'\Gamma(1)\beta_{\perp} \\ B &= (\alpha'\alpha)^{-1}\alpha'\Gamma(1)\beta_{\perp}(\alpha'_{\perp}\Gamma(1)\beta_{\perp})^{-1}. \end{aligned}$$

Luego

$$A\alpha' + B\alpha'_{\perp} = (\alpha'\alpha)^{-1}\alpha'(-I_k + \Gamma(1)\beta_{\perp}(\alpha'_{\perp}\Gamma(1)\beta_{\perp})^{-1}).$$

Por lo tanto, la matriz U dada en (3.46) es igual a:

$$U = \begin{pmatrix} (\alpha' \alpha)^{-1} \alpha' (-I_k + \Gamma(1) \beta_{\perp} (\alpha'_{\perp} \Gamma(1) \beta_{\perp})^{-1}) \\ (\alpha'_{\perp} \Gamma(1) \beta_{\perp})^{-1} \end{pmatrix}.$$

Retomando (3.45),

$$\begin{aligned} \Psi &= \beta_{\perp} [0 : I_{k-r}] [-\alpha : \Gamma(1) \beta_{\perp}]^{-1} \\ &= \beta_{\perp} [0 : I_{k-r}] \begin{pmatrix} A \alpha' + B \alpha'_{\perp} \\ C \alpha' + D \alpha'_{\perp} \end{pmatrix} \\ &= \beta_{\perp} I_{k-r} (C \alpha' + D \alpha'_{\perp}) \\ &= \beta_{\perp} (\alpha'_{\perp} \Gamma(1) \beta_{\perp})^{-1} \alpha'_{\perp}. \end{aligned} \quad (3.56)$$

Paso 6. Concluimos de (3.41), (3.44) y (3.56) que

$$\begin{aligned} y_t &= \underbrace{\beta_{\perp} (\alpha'_{\perp} \Gamma(1) \beta_{\perp})^{-1} \alpha'_{\perp}}_{\Psi} \sum_{j=1}^t \tilde{u}_j + \underbrace{(\beta_{\perp} \bar{\beta}'_{\perp} \Theta^*(L) + \bar{\beta} \beta' \Theta(L))}_{\Psi^*(L)} \tilde{u}_t \\ &\quad + \underbrace{(I_k - \bar{\beta} \beta') y_0 + (\Psi - \beta_{\perp} \bar{\beta}'_{\perp} \Theta_0) \tilde{u}_0}_{y_0^*} \end{aligned} \quad (3.57)$$

$$= \Psi \sum_{j=1}^t \tilde{u}_j + \Psi^*(L) \tilde{u}_t + \underbrace{(I_k - \bar{\beta} \beta') y_0 + \beta_{\perp} \bar{\beta}'_{\perp} \Theta^*(L) \tilde{u}_0}_{y_0^*}. \quad (3.58)$$

Luego, recordando que $\tilde{u}_t = u_t + \phi d_t$, tenemos que

$$\begin{aligned} y_t &= \underbrace{\beta_{\perp} (\alpha'_{\perp} \Gamma(1) \beta_{\perp})^{-1} \alpha'_{\perp}}_{\Psi} \sum_{j=1}^t (u_j + \phi d_j) + \underbrace{(\beta_{\perp} \bar{\beta}'_{\perp} \Theta^*(L) + \bar{\beta} \beta' \Theta(L))}_{\Psi^*(L)} (u_t + \phi d_t) \\ &\quad + \underbrace{(I_k - \bar{\beta} \beta') y_0 + (\Psi - \beta_{\perp} \bar{\beta}'_{\perp} \Theta_0) (u_0 + \phi d_0)}_{y_0^*} \\ &= \Psi \sum_{j=1}^t (u_j + \phi d_j) + \Psi^*(L) (u_t + \phi d_t) + \underbrace{(I_k - \bar{\beta} \beta') y_0 + \beta_{\perp} \bar{\beta}'_{\perp} \Theta^*(L) (u_0 + \phi d_0)}_{y_0^*}. \end{aligned} \quad (3.59)$$

Por último, es directo que $\beta' y_0^* = 0$ pues,

$$\begin{aligned} \beta' y_0^* &= \beta' [(I_k - \bar{\beta} \beta') y_0 + \beta_{\perp} \bar{\beta}'_{\perp} \Theta^*(L) (u_0 + \phi d_0)] \\ &= \beta' y_0 - \beta' \bar{\beta} \beta' y_0 + \beta' \beta_{\perp} \bar{\beta}'_{\perp} \Theta^*(L) (u_0 + \phi d_0) \\ &= \beta' y_0 - I_K \beta' y_0 = 0. \end{aligned}$$

□

El Teorema 3.29 es de importancia fundamental porque descompone el proceso y_t en componentes $I(1)$ e $I(0)$ que son útiles, por ejemplo, cuando se derivan las propiedades asintóticas de los estimadores de los parámetros. Además, como mencionamos previamente, precisa bajo qué condiciones el proceso y_t es conducido por $K - r$ componentes $I(1)$ y r componentes $I(0)$.

Observación. Debido al hecho que $B(z)$ es un polinomio matricial con $|B(z)| \neq 0$ para todo z tal que $|z| > 1$, $B(z)$ es invertible allí. Luego, existe $\Theta(z)$ polinomio matricial tal que $B(z)\Theta(z) = I$ y ambos polinomios pueden expresarse como serie de potencias con radio de convergencia $R = 1 + \delta$ para algún $\delta > 0$. Más aún, por la Proposición C.9 de Lütkepohl [43] y suponiendo que $\Theta(z) = \sum_{j=0}^{\infty} \theta_j z^j$, tenemos que el operador $\Theta(L) = \sum_{j=0}^{\infty} \theta_j L^j$ está bien definido y es el operador inverso de $B(L)$.

Por otro lado, definimos el operador $\Theta^*(L) = \sum_{j=0}^{\infty} \theta_j^* L^j$ con $\theta_j^* \doteq -\sum_{k=j+1}^{\infty} \theta_k$. Puesto que la serie $\Theta(1)$ converge, entonces los θ_j^* están bien definidos para todo $j \geq 0$. Veamos que $\Theta^*(L)$ está bien definido.

Sabemos que $\Theta'(z)$ está bien definido en $z = 1$ pues el radio de convergencia de $\Theta(z)$ es mayor que 1. Luego, dado que

$$\Theta'(1) = \sum_{j \geq 1} j \theta_j = \sum_{k \geq 0} \sum_{j \geq k+1} \theta_j = -\sum_{k \geq 0} \theta_k^* = -\Theta^*(1),$$

resulta que la serie $\Theta^*(1)$ es absolutamente convergente. Luego, por proposición C.9 de (Lütkepohl [43]), $\Theta^*(L)$ está bien definido.

Por otro lado, es importante notar que $\Theta(z) = \Theta(1) + (1 - z)\Theta^*(z)$. En particular,

$$\Theta(0) = \theta_0 = \sum_{k=0}^{\infty} \theta_k - \sum_{k=1}^{\infty} \theta_k = \Theta(1) + \theta_0^* = \Theta(1) + \Theta^*(0).$$

Veamos esto en general:

$$\begin{aligned} \Theta(z) &= \Theta(1) + \Theta(z) - \Theta(1) \\ &= \Theta(1) + (1 - z) \frac{\Theta(z) - \Theta(1)}{1 - z}. \end{aligned}$$

Ahora,

$$\begin{aligned} \Theta(z) - \Theta(1) &= \sum_{j \geq 0} \theta_j z^j - \Theta(1) \\ \frac{1}{1 - z} &= \sum_{j \geq 0} z^j = \sum_{j \geq 0} a_j z^j \quad a_j = 1 \quad \forall j \end{aligned}$$

Luego, para $|z| < 1$ se tiene que:

$$\begin{aligned}
(\Theta(z) - \Theta(1)) \frac{1}{1-z} &= \left(\sum_{j \geq 0} \Theta_j z^j - \Theta(1) \right) \sum_{j \geq 0} a_j z^j \\
&= \sum_{j \geq 0} \Theta_j z^j \sum_{j \geq 0} a_j z^j - \Theta(1) \sum_{j \geq 0} a_j z^j \\
&= \sum_{k \geq 0} \sum_{j=0}^k \Theta_j a_{k-j} z^j z^{k-j} - \Theta(1) \sum_{j \geq 0} a_j z^j \\
&= \sum_{k \geq 0} z^k \sum_{j=0}^k \Theta_j a_{k-j} - \sum_{k \geq 0} \Theta(1) a_k z^k.
\end{aligned}$$

De esta manera, tenemos que $(\Theta(z) - \Theta(1)) \frac{1}{1-z} = \sum_{k \geq 0} \Theta_k^* z^k$ donde los coeficientes Θ_k^* son

$$\Theta_k^* = \sum_{j=0}^k \Theta_j a_{k-j} - \Theta(1) a_k = \sum_{j=0}^k \Theta_j - \Theta(1) = - \sum_{j \geq k+1} \Theta_j.$$

Concluimos que

$$\frac{\Theta(z) - \Theta(1)}{1-z} = \Theta^*(z), \quad \text{para } |z| < 1$$

Luego, $\Theta^*(z)$ está bien definido en $z = 1$ pues:

$$\Theta^*(1) = \lim_{z \rightarrow 1} \Theta^*(z) = - \lim_{z \rightarrow 1} \frac{\Theta(z) - \Theta(1)}{z-1} = -\Theta'(1) = - \sum_{j=0}^{\infty} j \theta_j.$$

De esta manera, utilizando resultados de análisis de series numéricas, esta última serie converge y $\Theta^*(L)$ es un operador bien definido por proposición C.9 de Lütkepohl [43].

3.3.4. Análisis del término determinista en la Representación de Granger

Asumamos el siguiente modelo de corrección de error:

$$\Delta y_t = \phi d_t + \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t, \quad (3.60)$$

donde y_t es un proceso VAR(p) de dimensión K e $I(1)$ y ϕd_t un término determinista. Además, Π tiene rango menor a K , Γ_j son matrices constantes, $1 \leq j \leq p-1$, y $u(t)$ es un ruido blanco con media 0.

Asumiremos las siguientes denominaciones en la ecuación (3.31) del Teorema de representación de Granger: el término determinista ϕd_t se denomina *término de drift o deriva* mientras que el término determinista en la expresión para y_t dado por $\Psi \sum_{j=1}^t \phi d_j + \Psi^*(L) \phi d_t$ se denomina *tendencia*.

Del hecho que $\beta' y_0^* = 0$ y $\beta' \Psi = \underbrace{\beta' \beta_\perp}_{=0} \left[\alpha'_\perp \left(I_K - \sum_{i=1}^{p-1} \Gamma_i \right) \beta_\perp \right]^{-1} \alpha'_\perp = 0$, la combinación lineal $\beta' y_t$ en (3.31) queda expresada de la siguiente manera:

$$\beta' y_t = \beta' \Psi^*(L)(u_t + \phi d_t) = \beta' \Psi^*(L)u_t + \beta' \Psi^*(L)\phi d_t. \quad (3.61)$$

Luego,

$$E(\beta' y_t) = \beta' \Psi^*(L)\phi d_t. \quad (3.62)$$

Por otro lado, sean $z_t \doteq \Psi^*(L)u_t$ y $w_t \doteq \Psi^*(L)\phi d_t$. Entonces, z_t es un proceso estacionario e:

$$y_t = \Psi \sum_{j=1}^t (u_j + \phi d_j) + z_t + w_t + y_0^*.$$

De esta manera,

$$\Delta y_t = \Psi(u_t + \phi d_t) + \Delta z_t + \Delta w_t. \quad (3.63)$$

Por lo tanto,

$$E(\Delta y_t) = \Psi \phi d_t + E(\Delta w_t) = \Psi \phi d_t + \Delta w_t. \quad (3.64)$$

Consideremos $\phi d_t \doteq \mu_t = \mu_0 + \mu_1 t$ y algunas especificaciones de μ_t que se han propuesto en la literatura junto con sus implicaciones. Para analizar esto, descomponemos los parámetros μ_i en las direcciones de α y α_\perp de la siguiente manera:

$$\mu_i = \alpha \rho'_i + \alpha_\perp \gamma_i \quad i = 0, 1. \quad (3.65)$$

Esta descomposición nos permite desarrollar diferentes casos puesto que en la ecuación (3.31), la matriz $\Psi = \beta_\perp \left[\alpha'_\perp \left(I_K - \sum_{i=1}^{p-1} \Gamma_i \right) \beta_\perp \right]^{-1} \alpha'_\perp = \beta_\perp [\alpha'_\perp \Gamma(1) \beta_\perp]^{-1} \alpha'_\perp$ y esto implica que $\Psi \alpha = 0$. Luego, podemos detallar 5 casos de relevancia:

- **Caso 1:** Sin restricciones. En este caso, observamos que la representación de y_t queda expresada de la siguiente manera:

$$\begin{aligned} y_t &= \Psi \sum_{j=1}^t u_j + \Psi \mu_0 \sum_{j=1}^t 1 + \Psi \mu_1 \sum_{j=1}^t j + z_t + w_t + y_0^* \\ &= \Psi \sum_{j=1}^t u_j + \Psi \left(\mu_0 + \frac{\mu_1}{2} \right) t + \Psi \mu_1 \frac{t^2}{2} + z_t + w_t + y_0^* \end{aligned}$$

Entonces, el proceso permite una tendencia cuadrática determinada por

$$\tau_2 = \Psi \mu_1 = \beta_\perp [\alpha'_\perp \Gamma(1) \beta_\perp]^{-1} \alpha'_\perp (\alpha \rho'_1 + \alpha_\perp \gamma_1) = \beta_\perp [\alpha'_\perp \Gamma(1) \beta_\perp]^{-1} \alpha'_\perp \alpha_\perp \gamma_1.$$

Sin embargo, la combinación lineal $\beta' y_t$ no tiene tendencia cuadrática puesto que $\beta' \Psi = 0$ como vimos anteriormente, aunque sí posee tendencia lineal. Además, de (3.62) y (3.64), obtenemos que:

$$E(\beta' y_t) = \beta' \Psi^*(L) \mu_0 + \beta' \Psi^*(L) \mu_1 t.$$

y

$$E(\Delta y_t) = \Psi \mu_0 + \Psi \mu_1 t + \Delta w_t.$$

Ambos valores esperados muestran un comportamiento lineal.

- Caso 2: $\gamma_1 = 0$ ó $\alpha_{\perp} \gamma_1 = 0$. En este caso, $\mu_t = \mu_0 + \alpha \rho_1' t$ y está caracterizado por ausencia de tendencia cuadrática ya que $\tau_2 = \Psi \mu_1 = \Psi \alpha \rho_1' = \beta_{\perp} [\alpha_{\perp}' \Gamma(1) \beta_{\perp}]^{-1} \alpha_{\perp}' \alpha_{\perp} \alpha \rho_1' = 0$. luego, y_t queda expresado de la siguiente manera:

$$\begin{aligned} y_t &= \Psi \sum_{j=1}^t u_j + \Psi(\mu_0 + \alpha \rho_1' t) + \tau_2 \frac{t^2}{2} + z_t + w_t + y_0^* \\ &= \Psi \sum_{j=1}^t u_j + \Psi \mu_0 t + z_t + w_t + y_0^*. \end{aligned} \quad (3.66)$$

Sin embargo, aún permite la posibilidad de una tendencia lineal en todos los componentes del proceso, una tendencia que no puede ser eliminada por las relaciones de cointegración β :

$$\beta' y_t = \beta' \Psi^*(L) u_t + \beta' \Psi^*(L) \mu_0 + \beta' \Psi^*(L) \alpha \rho_1' t,$$

tal que

$$E(\beta' y_t) = \beta' \Psi^*(L) \mu_0 + \beta' \Psi^*(L) \alpha \rho_1' t.$$

Luego, el proceso βy_t es estacionario con tendencia. Por último,

$$E(\Delta y_t) = \Psi \mu_0 + \Psi \mu_1 t + \Delta w_t = \Psi \mu_0 + \Delta w_t.$$

Aquí sí observamos una media constante.

- Caso 3: $\mu_1 = 0$. En este modelo, $\mu_t = \mu_0$, tal que el proceso y_t tiene la forma (3.66), es decir, tiene una tendencia lineal dada por el coeficiente $\Psi \mu_0$, pero ésta puede ser eliminada por las relaciones de cointegración β , y el proceso contiene componentes estacionarios sin tendencia. Por lo tanto, el modelo permite una tendencia lineal en cada variable pero no en las relaciones de cointegración:

$$\beta' y_t = \beta' \Psi^*(L) u_t + \beta' \Psi^*(L) \mu_0.$$

Además,

$$E(\beta' y_t) = \beta' \Psi^*(L) \mu_0$$

y

$$E(\Delta y_t) = \Psi \mu_0 + \Delta w_t = \Psi \mu_0.$$

En la última igualdad usamos el hecho que $w_t = \Psi^*(L)\phi d_t = \Psi^*(L)\mu'_0$. Entonces, $\Delta w_t = 0$ pues asumimos que $La_0 = a_0$ para toda constante a_0 .

- **Caso 4:** $\mu_1 = 0$ y $\gamma_0 = 0$ ó $\alpha_\perp \gamma'_0 = 0$. Aquí no hay ninguna tendencia debido a que $\Psi \alpha \rho'_0 = 0$ con lo cual:

$$\begin{aligned} y_t &= \Psi \sum_{j=1}^t u_j + \Psi \mu_0 t + z_t + w_t + y_0^* \\ &= \Psi \sum_{j=1}^t u_j + \Psi \alpha \rho'_0 t + z_t + w_t + y_0^* \\ &= \Psi \sum_{j=1}^t u_j + z_t + w_t + y_0^*. \end{aligned}$$

Sin embargo, aún se permite un término constante en las relaciones de cointegración:

$$\begin{aligned} \beta' y_t &= \beta' \Psi^*(L) u_t + \beta' \Psi^*(L) \mu_0 \\ &= \beta' \Psi^*(L) u_t + \beta' \Psi^*(L) \alpha \rho'_0. \end{aligned}$$

Luego,

$$E(\beta' y_t) = \beta' \Psi^*(L) \alpha \rho'_0$$

y

$$E(\Delta y_t) = \Psi \alpha \rho'_0 + \Delta w_t = 0.$$

En la última igualdad usamos el hecho que $\Psi \alpha = 0$ y que $w_t = \Psi^*(L)\phi d_t = \Psi^*(L)\alpha \rho'_0$ lo que implica $\Delta w_t = 0$.

Consideremos ahora el caso particular que $\mu_0 = -\alpha \beta' k$ para algún $k \in \mathbb{R}^K$. Luego, $\mu_0 = \alpha \rho'_0$ pues $\alpha'_\perp \mu_0 = 0$.

De esta manera, la ecuación (3.60) para y_t queda expresada de la siguiente manera:

$$\begin{aligned} \Delta y_t &= -\alpha \beta' k + \alpha \beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t \\ &= \alpha \beta' (y_{t-1} - k) + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t. \end{aligned}$$

e

$$y_t = \Psi \sum_{j=1}^t u_j + z_t - \Psi^*(L) \alpha \beta' k + y_0^*$$

tal que $y_t = x_t + \tilde{k}$ con x_t representado por un VECM sin drift.

Dicha formulación da una interpretación fácil de k como el nivel de y_t , pero hay que tener en cuenta que el parámetro k no se identifica ya que sólo $\beta'k$ entra en el modelo. Esto es razonable ya que para el componente no estacionario, $\beta'_\perp y_t$, por ejemplo, el rol del nivel del proceso se toma por el valor inicial $\beta'_\perp y_0$.

- **Caso 5:** $\mu_0 = \mu_1 = 0$. Este modelo no permite término constante, lo que significa que todas las combinaciones lineales estacionarias tendrán media nula. Esto se ve:

$$y_t = \Psi \sum_{j=1}^t u_j + z_t + y_0^*,$$

$$\beta' y_t = \beta' \Psi^*(L) u_t$$

tal que

$$E(\beta' y_t) = 0$$

y

$$E(\Delta y_t) = 0.$$

3.4. Estimación del modelo de corrección de error

Siguiendo el libro de Lütkepohl [43] detallaremos la teoría correspondiente a la estimación por máxima verosimilitud de los modelos VECM. Nos focalizaremos en esta técnica debido a que las funciones ya implementadas en Python para estimación de VECM utilizan esta metodología.

Primero consideremos un modelo sin término determinístico.

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t \quad (3.67)$$

donde y_t es un proceso de dimensión K , $rk(\Pi) = r$ con $0 < r < K$ tal que $\Pi = \alpha\beta'$, donde α y β son matrices $K \times r$ con $rk(\alpha) = rk(\beta) = r$. Como siempre, Γ_j con $j = 1, \dots, p-1$ son matrices de parámetros $K \times K$ y $u_t \sim N(0, \Sigma_u)$ es ruido blanco con distribución normal estándar.

Además, y_t se supone un proceso $I(1)$ tal que

$$\alpha'_\perp \left(I_K - \sum_{i=1}^{p-1} \Gamma_i \right) \beta_\perp$$

es no singular (ver demostración del *Teorema de Representación de Granger*, paso 5, 3.53).

En lo que sigue, asumiremos que estas condiciones se cumplen cuando se considera el VECM (3.67). Además, a los fines de estimación, supondremos que se dispone de una muestra y_1, \dots, y_T así como de los valores de previos necesarios.

Luego, es conveniente escribir el VECM (3.67), para $t = 1, \dots, T$, en notación matricial como

$$\Delta Y = \Pi Y_{-1} + \Sigma \Delta X + U,$$

donde

$$\Delta Y \doteq [\Delta y_1, \dots, \Delta y_T],$$

$$Y_{-1} \doteq [y_0, \dots, y_{T-1}],$$

$$\Sigma \doteq [\Sigma_1, \dots, \Sigma_{p-1}],$$

$$\Delta X \doteq [\Delta X_0, \dots, \Delta X_{T-1}], \quad \text{con } \Delta X_{t-1} \doteq \begin{pmatrix} \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p+1} \end{pmatrix}$$

y

$$U \doteq [u_1, \dots, u_T].$$

3.4.1. Estimación ML

Si el proceso y_t es Gaussiano o, equivalentemente $u_t \sim N(0, \Sigma_u)$, el VECM (3.67) se puede estimar por máxima verosimilitud (ML) teniendo en cuenta la restricción de rango para $\Pi = \alpha\beta'$ (ver Johansen [35]). La función de máxima verosimilitud para una muestra de tamaño T es

$$\begin{aligned} \ln(l(\alpha, \beta, \Gamma, \Sigma_u)) &= -\frac{KT}{2} \ln(2\pi) - \frac{T}{2} \ln(|\Sigma_u|) \\ &\quad - \frac{1}{2} \text{tr}[(\Delta Y - \alpha\beta'Y_{-1} - \Sigma\Delta X)' \Sigma_u^{-1} (\Delta Y - \alpha\beta'Y_{-1} - \Sigma\Delta X)] \end{aligned} \quad (3.68)$$

A continuación, discutiremos el cálculo de los estimadores y luego consideraremos sus propiedades asintóticas.

Para la estimación ML no asumimos β normalizado. Sólo haremos la suposición de $rk(\Pi) = r$ lo que implica que la matriz se puede representar como $\Pi = \alpha\beta'$, donde α y β son $K \times r$ con $rk(\alpha) = rk(\beta) = r$. En la siguiente proposición se expondrán los estimadores de ML.

Proposición 3.4.1. (Estimadores ML de un VECM)

Sean $M \doteq I_T - \Delta X'(\Delta X \Delta X')^{-1} \Delta X$, $R_0 \doteq \Delta Y M$ y $R_1 \doteq Y_{-1} M$ y definimos

$$S_{ij} \doteq R_i R_j' / T \quad i = 0, 1$$

$$\lambda_1 \geq \dots \geq \lambda_K \quad \text{los autovalores de } S_{11}^{-1/2} S_{10} S_{00}^{-1} S_{01} S_{11}^{-1/2},$$

y

ν_1, \dots, ν_K los correspondientes autovectores ortonormales.

La función de máxima verosimilitud en (3.68) es maximizada por:

$$\begin{aligned}\beta &= \tilde{\beta} \doteq [\nu_1, \dots, \nu_r]' S_{11}^{-1/2}, \\ \alpha &= \tilde{\alpha} \doteq \Delta Y M Y_{-p}' \tilde{\beta} (\tilde{\beta}' Y_{-1} M Y_{-1}' \tilde{\beta})^{-1} = S_{01} \tilde{\beta} (\tilde{\beta}' S_{11} \tilde{\beta})^{-1}, \\ \Gamma &= \tilde{\Gamma} \doteq (\Delta Y - \tilde{\alpha} \tilde{\beta}' Y_{-1}) \Delta X' (\Delta X \Delta X')^{-1}, \\ \Sigma_u &= \tilde{\Sigma}_u \doteq (\Delta Y - \tilde{\alpha} \tilde{\beta}' Y_{-1} - \tilde{\Gamma} \Delta X) (\Delta Y - \tilde{\alpha} \tilde{\beta}' Y_{-1} - \tilde{\Gamma} \Delta X)' / T.\end{aligned}$$

El máximo es

$$\text{máx} \ln(l) = -\frac{KT}{2} \ln(2\pi) - \frac{T}{2} \left[\ln |S_{00}| + \sum_{i=1}^r \ln(1 - \lambda_i) \right] - \frac{KT}{2}$$

Demostración: Ver página 295 Lütkepohl [43].

Las soluciones $\tilde{\alpha}$ y $\tilde{\beta}$ del problema de optimización dado en la proposición previa no son únicas. Para cualquier matriz Q no singular $r \times r$, $\tilde{\alpha}Q^{-1}$ y $\tilde{\beta}Q'$ representan otro conjunto de estimadores de ML para α y β . Sin embargo, la proposición muestra que existen expresiones explícitas para los estimadores de ML. Si $r = K$, la proposición sigue siendo válida. Además, los estimadores de ML para los niveles de la representación VAR correspondientes al VECM (3.67) observando la restricción de rango se pueden obtener de manera sencilla a través de las siguientes relaciones:

$$\begin{aligned}A_1 &= \Pi + I_K + \Gamma_1, \\ A_i &= \Gamma_i - \Gamma_{i-1}, \quad i = 2, \dots, p-1, \\ A_p &= -\Gamma_{p-1}.\end{aligned}$$

Capítulo 4

Modelo de cointegración multifactorial

4.1. Introducción

En este trabajo, nuestro objetivo es plantear un modelo matemático que sea útil para explicar los movimientos de la curva de Swap Spread a través de los principales drivers o fundamentales económicos que afectan tal curva de crédito. Además, este modelo debe satisfacer los siguientes puntos:

- Describir los movimientos de manera estructural más que intentar entender los cambios en cada punto de la curva.
- No asumir estacionariedad en los factores y drivers.
- Parsimonioso y sencillo de estimar.
- Intentar buscar una relación analítica entre factores y drivers.

Debido a que el comercio de bonos y futuros sobre tasas de interés que manejaremos en este trabajo se lleva a cabo en mercados profundos y bien organizados, resulta atractivo imponer restricciones teóricas que eliminen las posibilidades de arbitraje libre de riesgo a través de los distintos plazos y en el tiempo.

Como mencionamos en el capítulo 2, Diebold and Li [18] proponen un modelo de dos etapas basado en el marco de Nelson and Siegel [47] para pronosticar la estructura a término de Estados Unidos que presentó mejores resultados que algunos modelos competidores a pesar de no imponer restricciones de no arbitraje. Sin embargo, Almeida and Vicente [1] (2008) muestran que la inclusión de condiciones de no arbitraje en un modelo latente mejora el ajuste fuera de muestra. Ang and Piazzesi [2] (2003) encuentran que un modelo afín con variables macroeconómicas supera al modelo VAR sin restricciones que contiene los mismos factores observables. También muestran que los modelos con macro factores predicen mejor que los modelos con sólo factores no observables. En la misma línea, Hördahl et al. [33] (2006) confirman

que el desempeño de predicción de un modelo con factores observables es superior a los modelos basados en factores latentes.

Lamentablemente, los modelos afines canónicos libres de arbitraje a menudo no tienen un buen desempeño empírico con respecto a series de tiempo, es decir, a lo referido a los ajustes dinámicos. Esto se observa, principalmente, cuando se estudia la predicción de las tasas futuras, un punto importante señalado por Duffee [22]. Tal ajuste dinámico es crucial para nuestro objetivo de relacionar la evolución de la curva de tasas en el tiempo a los movimientos en los drivers o variables macroeconómicas. Además, y de manera crucial, la estimación de estos modelos es conocido por ser problemático, en gran parte debido a la existencia de numerosas funciones de máxima verosimilitud que tienen esencialmente idénticos ajustes a los datos pero implicaciones muy diferentes sobre el comportamiento económico.

Por otro lado, buscamos un modelo que incorpore explícitamente variables económicas fundamentales. Si bien en los últimos años se han desarrollado modelos que incorporan variables fundamentales, como es el caso del trabajo de Kobor et al. [36] por el Banco Mundial, el modelo obtenido arroja resultados relacionados con 3 puntos particulares de la curva de Swap Spread correspondiente a los vencimientos a 2, 5 y 10 años. Sin embargo, un desafío de gran interés es intentar captar los movimientos de una manera más global a lo largo de la curva. Además, buscamos brindarle a nuestro modelo un marco de mayor formalidad matemática. Por otro lado, es importante que continúe siendo parsimonioso, simple y que incorpore los drivers directamente al modelo, lo cual se aleja de trabajos como el de Feldhütter and Lando [27] donde si bien se presenta una estructura formal e imposiciones sobre libre arbitraje, también se plantea un modelo de tipo afín complejo de 6 factores y donde no se incluyen los drivers directamente al modelo.

En el artículo de Diebold et. al [20] se busca caracterizar las relaciones entre $\{L_t\}$, $\{S_t\}$ y $\{C_t\}$ y la macroeconomía a través de un modelo VAR. Los autores incluyen tres variables clave como medidas de la economía: la utilización de la capacidad de fabricación, la tasa de fondos federales y la inflación anual de los precios. Luego, estiman una extensión del modelo de Nelson y Siegel para la curva de tasas del Tesoro de EE.UU. considerando 6 series de tiempo asociadas a nivel, pendiente y curvatura junto con las tres variables macroeconómicas. Es decir, obtienen $\{L_t\}$, $\{S_t\}$, $\{C_t\}$, $\{CU_t\}$, $\{FFR_t\}$ y $\{INFL_t\}$ a partir de Nelson y Siegel dinámico y lo plantean como un VAR. Sin embargo, asumen un modelo estacionario tanto para los factores como para las variables económicas a pesar de que se concluye lo contrario a través de diversos tests estadísticos como el tests de Dickey-Fuller (DF) o Dickey-Fuller aumentado (ADF) entre otros.

Por lo tanto, para llevar a cabo nuestro objetivo de capturar las dinámicas de la curva de manera estructural, primero analizaremos este último modelo VAR aplicado a la curva de Swap Spread junto con

los drivers que la conducen. Luego, descartaremos la hipótesis de estacionariedad para incorporar las relaciones de cointegración entre factores y drivers asumiendo un VECM para los mismos.

Las variables macroeconómicas o drivers a considerar se obtendrán de las planteadas en el artículo del Banco Mundial de Kobor et al. [36]: “What Determines U.S. Swap Spreads?”.

Como mencionamos al comienzo de este trabajo, el modelo que proponen en Diebold et al. [20] tiene la flexibilidad necesaria para adaptarse a la forma cambiante de la curva de rendimiento además de ser parsimonioso y fácil de estimar. Mostraremos que la incorporación de integrabilidad y la suposición de cointegración entre las variables no sólo mantiene la parsimonia sino que mejora el desempeño del modelo original. Además, permite determinar relaciones lineales explícitas a corto y largo plazo entre los factores que definen la curva y los drivers económicos. Por último, le dará rigurosidad matemática al modelo ya que utilizaremos la teoría desarrollada por Johansen y que describimos en el Capítulo 3.

Existen varias maneras para determinar la eficiencia de un modelo con respecto a otro a través de estimadores obtenidos con cada uno.

Una propiedad que se puede estudiar de cada estimador es la ubicación del valor esperado de su distribución. Esto es, el sesgo del estimador. Sin embargo, ser insesgado no es una propiedad deseable ya que es muy probable que proporcione una estimación incorrecta de una muestra dada. Además, un estimador insesgado puede tener una varianza extremadamente grande. Luego, es difícil determinar la utilidad de un estimador insesgado con una gran varianza. Restringir la definición de eficiencia a estimadores insesgados parece arbitrario y quizás poco útil. Puede ser que algunos estimadores sesgados con variaciones más pequeñas sean más útiles en la estimación. Un enfoque simple es comparar estimadores basados en su Error Cuadrático Medio (MSE):

$$\begin{aligned} MSE(\hat{\theta}) &\doteq E((\hat{\theta} - \theta)^2) \\ &= Var(\hat{\theta}) + [Sesgo(\hat{\theta})]^2 \quad \text{con } Sesgo(\hat{\theta}) \doteq E(\hat{\theta}) - \theta, \end{aligned}$$

donde $\hat{\theta}$ es un estimador de θ , un parámetro de población desconocido.

Por lo tanto, en este trabajo utilizaremos el criterio del Error Cuadrático Medio (MSE) para analizar estimadores obtenidos de predicciones in-sample. De esta manera, podremos comparar la eficiencia del modelo VAR original y nuestro modelo VECM en función de su desempeño predictivo.

La forma en que se llevará a cabo esta comparación está detallado en las Secciones 4.4 y 4.6 del presente capítulo.

4.2. Especificación del modelo VAR aplicado a DNS y drivers (VAR DNS)

En esta sección comenzaremos desarrollando la primera parte de nuestro modelo. Plantearemos la estructura de factor propuesta por Nelson and Siegel [47] en su forma dinámica para nuestra curva de Swap Spread de manera tal que dicha curva quede expresada como función de 3 factores latentes asociados al nivel, pendiente y curvatura (Diebold and Li [18]). Luego, incorporaremos las variables macroeconómicas o drivers que asumimos afectan a dicha curva (Diebold et al. [20]).

Como desarrollamos anteriormente en la Sección 2.3, si $T \in \mathbb{N}$ es el horizonte temporal, podemos extraer 3 factores latentes L_t , S_t y C_t en cada instante $t \in \{1, 2, \dots, T\}$. Estos factores se obtienen a través de la muestra de curva de tasas en cada t utilizando el modelo de Nelson y Siegel dinámico como se plantea en las ecuaciones (2.3) y en su forma matricial, ecuación (2.4):

$$\begin{pmatrix} \tilde{y}(t, \tau_1) \\ \tilde{y}(t, \tau_2) \\ \vdots \\ \tilde{y}(t, \tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{pmatrix} \begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix} + \begin{pmatrix} \epsilon_t(\tau_1) \\ \epsilon_t(\tau_2) \\ \vdots \\ \epsilon_t(\tau_N) \end{pmatrix}, \quad (4.1)$$

donde $\tilde{y}(t, \tau_1), \tilde{y}(t, \tau_2), \dots, \tilde{y}(t, \tau_N)$, con $N \in \mathbb{N}_{\geq 3}$ son las tasas de interés observadas en cada tiempo $t \in \{1, 2, \dots, T\}$ tal que $\tau_i = T_i - t$ y $0 < T_1 < T_2 < \dots < T_N$ son los vencimientos.

En este trabajo asumiremos un modelo vectorial autorregresivo como en Diebold et al. [20] que es útil para poder caracterizar las relaciones de cointegración entre los factores y los drivers. Escribiremos X_t para denotar un vector con uno, dos o tres factores. Sea $X_t \in \mathbb{R}^k$ un vector de k variables de series de tiempo, cada una representando un factor L_t, S_t o C_t , $1 \leq k \leq 3$, y sea $D_t \in \mathbb{R}^M$, $M > 0$, un vector de drivers. Entonces, si $k = 3$ tenemos:

$$X_t = \begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix}, \quad D_t = \begin{pmatrix} D_t^1 \\ D_t^2 \\ \vdots \\ D_t^M \end{pmatrix} \quad t = 1, 2, \dots, T-1, T. \quad (4.2)$$

Luego, asignamos un modelo vectorial autorregresivo VAR(p) a la serie vectorial extendida $\tilde{X}_t \in \mathbb{R}^{k+M}$ con

$$\tilde{X}_t = \begin{pmatrix} X_t \\ D_t \end{pmatrix} \quad t = 1, 2, \dots, T-1, T.$$

4.3. ESPECIFICACIÓN DEL MODELO VECM APLICADO A DNS Y DRIVERS (VECM DNS)89

En general, el modelo viene dado por:

$$\tilde{X}_t = A_1 \tilde{X}_{t-1} + A_2 \tilde{X}_{t-2} + \dots + A_p \tilde{X}_{t-p} + C + u_t, \quad (4.3)$$

con $A_1, \dots, A_p \in \mathbb{R}^{(k+M) \times (k+M)}$, $C \in \mathbb{R}^{k+M}$ el intercepto y $u_t \sim N(0, \Sigma_t)$ una serie vectorial de ruido blanco. La notación Σ_t representa la matriz de covarianza, no necesariamente diagonal. Por simplicidad, en este trabajo asumiremos \tilde{X}_t un proceso VAR(p) con $p = 1$ o $p = 2$.

En este punto observamos que el proceso VAR constituye el modelo sin cointegración entre factores y drivers que compararemos con el modelo VECM.

4.3. Especificación del modelo VECM aplicado a DNS y drivers (VECM DNS)

En el trabajo del Banco Mundial [36], los autores plantean un modelo que busca explicar los movimientos de los Swap Spreads destacando las fuerzas económicas que los conducen. Esto les permite conseguir una interpretación económica más significativa sobre los movimientos de tales Spreads.

Los autores revisan la teoría y los estudios empíricos pasados sobre los U.S. Swap Spreads y aplican una metodología Corrección de Error (EC) basada en el concepto de cointegración estimando un Modelo de Corrección de Error (ECM) para los vencimientos de 2, 5, y 10 años para el período de 1994 a 2004. Ellos encuentran que los U.S. dollar Swap Spreads y la oferta de bonos del Tesoro de EE.UU. están cointegrados, sugiriendo que la oferta del Tesoro es un factor determinante en un horizonte de largo plazo. Los autores estiman un Modelo de Corrección de Error que integra esta relación a largo plazo con la influencia de cuatro factores determinantes de corto plazo: el spread AA, la tasa repo, la diferencia entre los rendimientos on-the-run y off-the-run, y la duración de los mortgage-backed securities. Para algunos de estos drivers, se considera además como un determinante sus diferencias o lags de algún orden conformando así 7 determinantes. En otro modelo, la variable explicativa repo se excluye y se reemplaza por las primeras diferencias de la pendiente de 2 a 10 años del Treasury tanto en sus valores actuales como en sus diferencias.

Como un resultado del trabajo de especificación y de selección de variables y lags descriptas en su trabajo, ellos estiman dos modelos ECM para cada una de los vencimientos: 2, 5 y 10 años utilizando datos cuatrimestrales. De esta manera,

$$\text{SwSpr}_t = \alpha + \beta \cdot \text{TsySup}_t + \epsilon_t$$

es la relación a largo plazo, donde SwSpr_t es la serie de tiempo correspondiente al Swap Spread de cada vencimiento, TsySup_t es la serie asociada a la oferta de Tesoro y ϵ_t es la serie de los errores.

El término de corrección de error está definido como el error de estimación de esta ecuación de mínimos cuadrados ordinarios, denotado por ECT_t

$$ECT_t = \text{SwSpr}_t - (\alpha + \beta \cdot \text{TsySup}_t).$$

ECT_t es una de las variables explicatorias en la segunda ecuación del modelo, la relación de corto plazo involucrando las primeras diferencias de otras variables no estacionarias

$$\Delta \text{SwSpr}_t = C + \beta_0 ECT_{t-1} + \beta_1 \Delta \text{SwSpr}_{t-1} + \sum_{i=2}^8 \beta_i \Delta \text{Det}_i,$$

donde cada Det_i es uno de los 7 determinantes que fueron identificados para ambos modelos.

Nuestro objetivo es aplicar la metodología de cointegración a las series de tiempo $\{L_t\}$, $\{S_t\}$ y $\{C_t\}$ extraídas de la curva de Swap Spread utilizando Nelson y Siegel dinámico junto con los drivers pero a través de un Modelo de Corrección de Error Vectorial (VECM) en lugar de un ECM.

Esto nos permitirá obtener una visión más global y estructurada de las dinámicas entre factores y drivers.

A continuación daremos la especificación del VECM con Drivers.

Sea M el número de drivers a considerar. Sean $D_t \in \mathbb{R}^M$ y $X_t \in \mathbb{R}^k$ con $1 \leq k \leq 3$ como en la ecuación (4.2), los procesos vectoriales dados por los drivers y los factores, respectivamente. También suponemos que cada uno de los componentes de X_t y D_t sigue un proceso $\text{AR}(p)$ estacionario o integrado de orden 1, es decir, $I(0)$ o $I(1)$, y que también existe una cointegración entre los componentes de X_t y D_t . La hipótesis de cointegración entre las variables significa que existen coeficientes a_{ji} con $j = 1, \dots, M$ y $i = 1, 2, 3$ (no todos cero), tales que

$$L_t - (a_{11}d_{1t} + a_{21}d_{2t} + \dots + a_{j1}d_{jt}), \quad 1 \leq j \leq M$$

$$S_t - (a_{12}d_{1t} + a_{22}d_{2t} + \dots + a_{j2}d_{jt}), \quad 1 \leq j \leq M$$

$$C_t - (a_{13}d_{1t} + a_{23}d_{2t} + \dots + a_{j3}d_{jt}), \quad 1 \leq j \leq M$$

son series de tiempo estacionarias.

Definamos la serie temporal vectorial \tilde{X}_t por:

$$\tilde{X}_t = \begin{pmatrix} X_t \\ D_t \end{pmatrix}.$$

Entonces

$$\underbrace{\begin{pmatrix} I_3 & : & -(a_{ji}) \end{pmatrix}}_{\beta'} \cdot \tilde{X}_t \tag{4.4}$$

es una serie de tiempo estacionaria. Supongamos que \tilde{X}_t sigue un modelo VAR(p) como en la ecuación (4.3):

$$\tilde{X}_t = A_1 \tilde{X}_{t-1} + A_2 \tilde{X}_{t-2} + \cdots + A_p \tilde{X}_{t-p} + C + u_t,$$

para algunas matrices constantes A_1, \dots, A_p , con A_p matriz distinta de cero, y u_t un ruido blanco.

Por otro lado, las primeras diferencias de \tilde{X}_t son estacionarias ya que sus componentes son $I(1)$. Entonces se pueden escribir como:

$$\Delta \tilde{X}_t = -(I - A_1 - A_2 - \cdots - A_p) \tilde{X}_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta \tilde{X}_{t-p+i} + C + u_t,$$

donde las matrices Γ_j se obtienen de las matrices A_i , $1 \leq j < p$, $1 \leq i \leq p$ (ver Sección 3.3.1).

Debido a que las primeras diferencias $\Delta \tilde{X}_t$ son estacionarias pero \tilde{X}_t no lo es, se sigue que la matriz $\Pi = -(I - A_1 - A_2 - \cdots - A_p)$ es no invertible. Luego, el rango r de Π satisface $r < M + k$ y Π se puede escribir como:

$$-(I - A_1 - A_2 - \cdots - A_p) = \Pi = \alpha \beta'$$

con $\alpha, \beta \in \mathbb{R}^{(M+k) \times r}$. En particular, β se puede elegir de una manera normalizada tal que $\beta_{ij} = \delta_{ij}$, para $1 \leq i, j \leq r$. Luego, el modelo queda expresado de la siguiente manera:

$$\Delta \tilde{X}_t = \alpha \beta' \tilde{X}_{t-1} + \Gamma_1 \Delta \tilde{X}_{t-1} + \cdots + \Gamma_{p-1} \Delta \tilde{X}_{t-p+1} + C + u_t, \quad \forall t : 0 \leq t \leq T. \quad (4.5)$$

Vamos a generalizar nuestro modelo, asumiendo que el proceso \tilde{X}_t puede presentar términos deterministas en el modelo de cointegración:

$$\Delta \tilde{X}_t = \phi d_t + \alpha \beta' \tilde{X}_{t-1} + \Gamma_1 \Delta \tilde{X}_{t-1} + \cdots + \Gamma_{p-1} \Delta \tilde{X}_{t-p+1} + u_t, \quad \forall t : 0 \leq t \leq T \quad (4.6)$$

- d_t es un término determinista que puede contener una constante, un término lineal, entre otros regresores fijos y no estocásticos.
- Se cumple que la matriz $\alpha \beta'$ es no invertible con rango r tal que $0 < r < M + K$. Es decir, existen r relaciones de cointegración entre factores y drivers.

Por lo tanto, tenemos dos modelos autorregresivos diferentes que representan la dinámica del vector de factores y drivers para las curvas de tasas Swap Spread. Se pretende mostrar que si las series están cointegradas, el modelo VECM tiene un mejor desempeño que el modelo VAR. La comparación entre los dos modelos se realizará analizando la eficiencia predictiva de los modelos (4.3) y (4.6) analizando la varianza de la media muestral para cada estimador con respecto al valor real de los factores. Esto se desarrolla en la siguiente sección.

4.4. Comparación entre los modelos VAR y VECM aplicados a DNS y drivers

Como mencionamos al comienzo del capítulo, hay varias maneras de determinar la eficiencia de un modelo con respecto a otro a través de estimadores obtenidos con cada uno.

En este trabajo, tomaremos un dato de la muestra de factores y drivers y utilizaremos todos los datos anteriores para ajustar los modelos VAR y VECM. Luego, podemos simular el último dato de la muestra y comparar que modelo lo estimó mejor. Para esto nos basaremos en los siguientes resultados y técnicas estadísticas:

- $\theta \doteq$ dato a estimar;
- $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_N : N$ simulaciones independientes de θ con uno de los modelos (VAR o VECM),
- $\hat{\theta} = \hat{\theta}(N) \doteq$ Media muestral de N simulaciones del dato a estimar;

donde la media muestral de N simulaciones se define como:

$$\hat{\theta}(N) \doteq \frac{(\hat{\theta}_1 + \dots + \hat{\theta}_N)}{N}.$$

Nota. Para nuestro estudio, cada simulación corresponderá al valor x_T de nuestra serie vectorial. Además, este último valor estará basado en $T - 1$ observaciones anteriores dadas por x_1, \dots, x_{T-1} que a su vez serán simulaciones de la serie temporal con cada modelo.

Las variables $\hat{\theta}_i, 1 \leq i \leq N$ tienen una distribución desconocida **con media μ y varianza σ** . Pero por la Ley Fuerte de los Grandes Números sabemos que $\hat{\theta}(N)$ converge en probabilidad (p.p.) a μ y que la varianza de $\hat{\theta}_N$ es igual a σ^2/N . Es decir:

$$\lim_{N \rightarrow \infty} \hat{\theta}(N) = \mu \text{ p.p.} \quad \text{y} \quad \text{Var}(\hat{\theta}(N)) = \frac{\sigma^2}{N}.$$

Asumiendo que $\text{VAR}(\hat{\theta}_i) < \infty$ se van a comparar los desvíos estándar de los estimadores para cada modelo y así evaluar su eficiencia.

Por otro lado, si el estimador $\hat{\theta}$ del parámetro θ fuera insesgado, es decir, si $\text{Sesgo}(\hat{\theta}) = \mu - \theta = 0 \Leftrightarrow E(\hat{\theta}_i) = \theta$ se tendría que

$$\lim_{N \rightarrow \infty} \hat{\theta}(N) = \theta \text{ p.p.}$$

Sin embargo, como comentamos previamente, ser insesgado no es necesariamente una propiedad deseada, además de que sería una suposición muy fuerte. Luego, asumiremos que el estimador tiene sesgo y lo podemos estimar de la siguiente manera:

$$Sesgo(\hat{\theta}_i) = E(\hat{\theta}_i) - \theta \approx \hat{\theta}(N) - \theta,$$

cuando N es suficientemente grande.

A continuación, formalizamos estas ideas para utilizarlas en nuestro trabajo.

Si consideramos el modelo VAR (4.3), la predicción del valor \tilde{X}_t dadas las observaciones hasta $t - 1$ está dada por el valor esperado condicional:

$$\begin{aligned} \hat{X}_{tVAR} &= E[\tilde{X}_t \mid \tilde{X}_s = x_s \text{ para } 0 \leq s \leq t - 1] \\ &= C + \sum_{i=1}^p A_i x_{t-i}. \end{aligned} \quad (4.7)$$

Aquí, \hat{X}_{tVAR} corresponde al parámetro θ previamente definido, y se obtiene a partir de las $t - 1$ observaciones anteriores aplicando el modelo VAR.

Por otro lado, si consideramos el modelo VECM en (4.6), la predicción del valor \tilde{X}_t dadas las observaciones hasta $t - 1$ está dada por el valor esperado condicional:

$$\begin{aligned} \hat{X}_{tVECM} &= E[\tilde{X}_t = \Delta\tilde{X}_t + \tilde{X}_{t-1} \mid X_s = x_s \text{ para } 0 \leq s \leq t - 1] \\ &= \phi d_t + (1 + \Pi)x_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta x_{t-p+i}. \end{aligned} \quad (4.8)$$

Aquí, \hat{X}_{tVECM} corresponde al parámetro θ previamente definido, y se obtiene a partir de las $t - 1$ observaciones anteriores aplicando el modelo VECM.

En este trabajo compararemos las estimaciones obtenidas por los modelos VAR y VECM respectivamente según los estimadores dados en (4.7) y (4.8).

En particular, analizaremos la predicción de los factores que intervienen en el modelo y que corresponden a una j -ésima componente del vector \tilde{X}_t .

Proponemos realizar este análisis mediante simulaciones de la serie de tiempo $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$ con cada modelo, y compararemos la eficiencia de los estimadores en la predicción de la componente $(\tilde{X}_{n+1})_j$ para un valor particular de j que representa uno de los factores. Para hacer esto, si el modelo autorregresivo es de orden p tomaremos el valor inicial $p - 1$ desde los datos muestrales y generaremos una serie de tamaño n de acuerdo con los parámetros de nuestro modelo. La componente j -ésima de la predicción $\hat{X}_{t,VAR}$ y $\hat{X}_{t,VECM}$ es luego un estimador de $(\tilde{X}_{n+1})_j$. Dado que la media muestral es un estimador consistente de la esperanza cuya varianza disminuye al aumentar el tamaño de la muestra, compararemos la varianza empírica del estimador de la media muestral del componente j -ésimo para cada uno de los modelos. Usaremos la siguiente definición de eficiencia.

Definición 4.4.1. [Eficiencia en la j -ésima componente] Sea $\delta > 0$ una tolerancia fija y sea $\hat{X}_{t,VAR}$ y $\hat{X}_{t,VECM}$ los estimadores del vector \tilde{X}_t como en las ecuaciones (4.7) y (4.8). Sean $(y_i)_j, 1 \leq i \leq N_1$ y

$(z_l)_j$, $1 \leq l \leq N_2$ las j -ésimas componentes de las estimaciones de \tilde{X}_t obtenidas de N_1 y N_2 simulaciones independientes de los estimadores $\hat{X}_{t,VAR}$ y $\hat{X}_{t,VECM}$ respectivamente, tal que:

$$\frac{1}{N_1 - 1} \sum_{i=1}^{N_1} ((y_i)_j - \bar{y})^2 < \delta^2, \quad \frac{1}{N_2 - 1} \sum_{l=1}^{N_2} ((z_l)_j - \bar{z})^2 < \delta^2$$

$$\bar{y} = \frac{1}{N_1} \sum_{i=1}^{N_1} (y_i)_j, \quad \bar{z} = \frac{1}{N_2} \sum_{l=1}^{N_2} (z_l)_j$$

y N_1 y N_2 son mínimos con esta propiedad. En este caso diremos que el modelo VAR es más eficiente que el modelo VECM (respectivamente el modelo VECM es más eficiente que el modelo VAR) para la estimación de la j -ésima componente de \tilde{X}_t si $N_1 < N_2$ (respectivamente si $N_2 < N_1$).

En otras palabras, diremos que un estimador es más eficiente que otro si requiere menos simulaciones para obtener una desviación estándar menor que δ .

De esta manera, los pasos para comparar los modelos VAR y VECM presentados anteriormente según la Definición 4.4.1 son los siguientes.

1. Consideramos la muestra de datos diarios de la curva de Swap Spread $\{y_t\}$ con vencimientos a 2, 3, 5, 7, 10 y 30 años correspondientes al período del 11 de noviembre de 2009 al 23 de octubre de 2015 obteniendo así un total de $T = 1500$ curvas.
2. Para esta serie de datos, se estiman a partir del modelo DNS las series individuales de nivel, pendiente y curvatura $\{L_t\}$, $\{S_t\}$ y $\{C_t\}$, $1 \leq t \leq T = 1500$. Esto se lleva a cabo utilizando cuadrados mínimos para un parámetro λ fijo.
3. Se analiza cointegración de estos factores con la serie de uno o más drivers.
4. Para cada modelo, los parámetros se estiman a partir de una submuestra de tamaño $T^* - 1$, $\{y_t\}_{t=1}^{T^*-1}$, donde la cointegración es observable.
5. Una vez estimados los parámetros del modelo simulamos una muestra de tamaño $T^* - 1$, tomando como datos iniciales los de la muestra original. Por ejemplo, para un $VAR(p)$, se tomarán y_1, y_2, \dots, y_{p-1} .
6. Predecimos el último paso con cada modelo utilizando la muestra generada de tamaño $T^* - 1$. Es decir, ajustamos a cada modelo la serie de tiempo generada y predecimos el dato correspondiente a T^* .
7. El procedimiento anterior se repite hasta que la varianza de la media muestral de los datos que predecimos sea menor que una tolerancia establecida de antemano.

8. Finalmente, compararemos el número necesario de simulaciones que cada modelo necesitó para alcanzar esa tolerancia. Concluimos que el modelo con menor número de simulaciones tiene un mejor desempeño y poder explicativo de la serie involucrada.

4.5. Estimación de los modelos

4.5.1. Datos y estimación del modelo VAR DNS

El primer punto de nuestro trabajo consistió en analizar las series de los factores de Nelson y Siegel obtenidas de los datos de yields de Swap Spread con vencimientos a 2, 3, 5, 7, 10 y 30 años. Para ello, consideramos los datos diarios correspondientes al período comprendido desde el 11 de Noviembre del 2009 hasta el 23 de Octubre del 2015 obteniendo así un total de $T = 1500$ yields para cada vencimiento. Los datos fueron obtenidos de Bloomberg y fueron brindados por la empresa Arfima Trading SL.

Las series de tiempo fueron estimadas utilizando mínimos cuadrados lineales, con lo cual fue necesario fijar el valor de λ . Basándonos en el resultado de la Nota B.1 que detalla la relación de λ con la curvatura de la función, escogimos dicho parámetro de la siguiente manera:

- Calculamos el número de extremos en cada tramo de la curva de Swap Spread para el total de datos y analizamos los cambios de pendiente.
- Observamos que en los tramos 2-3-5, 5-7-10 y 7-10-30 es donde se encuentran la mayoría de los extremos aunque de orden similar. Sin embargo, en el tramo 7-10-30 los cambios de curvatura que dan lugar a los extremos son más pronunciados. Además, es el tramo más largo de la curva. Luego, decidimos escoger $\tau_0 = 8$ años = 96 meses.
- Utilizando el punto de máximo de la tercer carga $x_0 \sim 1.7933$, calculamos el valor de $\lambda = x_0/\tau_0$.

Así, el valor de λ a considerar es $\lambda = 1.7933/96 = 0.01868$.

Los plazos considerados son equivalentes en meses a 24, 36, 60, 84, 120 y 360. Si bien nosotros vamos a considerar la curva de Swap Spread con vencimientos mínimos de 2 años, uno puede obtener tasas swap spread con plazos de vencimiento desde 1 mes (como mencionamos en la Sección 1.3). Por este motivo y porque se obtiene un mejor ajuste de la curva, decidimos expresar los vencimientos en meses para el cálculo de λ .

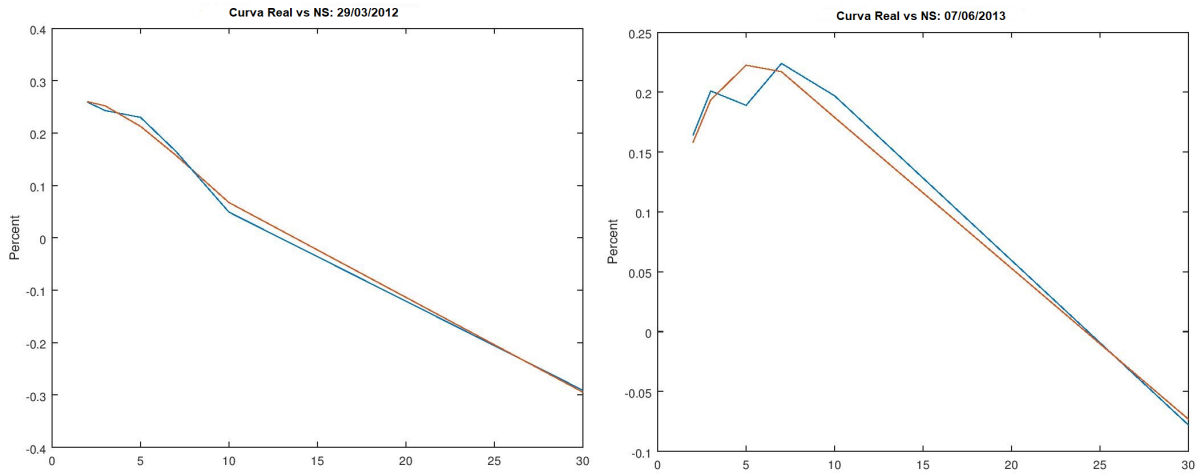
A continuación vemos algunos estadísticos que sustentan nuestra visión: Asumamos $\lambda_1 = 1.7933/96 = 0.01868$ y $\lambda_2 = 1.7933/8 = 0.22416$. Si estimamos los factores para el total de datos, estamos realizando 1500 estimaciones por mínimos cuadrados lineales para cada λ . Luego, si el estadístico R^2 para cada i -ésima estimación y para cada λ lo denotamos por $R_{i,1}^2$ y $R_{i,2}^2$, obtenemos:

- $\min_i R_{i,1}^2 = 0.6448 > \min_i R_{i,2}^2 = 0.1486$.

- $\max_i R_{i,1}^2 = 0.9998 > \max_i R_{i,2}^2 = 0.9438$.
- $Var(R_1^2) = 0.0028 < Var(R_2^2) = 0.0264$.

En los siguientes gráficos (ver Figura 4.1), observamos el resultado arrojado por el método de Nelson y Siegel en dos puntos de la muestra junto con la curva real.

Figura 4.1: Las curvas de color azul corresponden a las observadas y las de color rojo a las estimadas con el método de Nelson y Siegel



Una vez obtenida la serie de tiempo vectorial $\{X_t\}_{t=1}^{1500}$ para el nivel, pendiente y curvatura, realizamos 3 estudios de simulación que implicaron la estimación del modelo VAR planteado en la fórmula (4.2). La estimación de dicho modelo para cada caso de estudio se realizó utilizando una estimación por mínimos cuadrados multivariado como se detalla en el capítulo 3 de Lütkepohl [43].

4.5.2. Análisis inicial y estimación del modelo VECM DNS

Al igual que en el modelo previo, las series de los factores se obtuvieron con la metodología de Nelson y Siegel a partir de los datos de yields de Swap Spread con vencimientos a 2, 3, 5, 7, 10 y 30 años y se consideraron los datos diarios correspondientes al período comprendido desde el 11 de Noviembre del 2009 hasta el 23 de Octubre del 2015 obteniendo así un total de $T = 1500$ yields para cada vencimiento. Los datos fueron obtenidos de Bloomberg y fueron brindados por la empresa Arfima Trading SL.

Las series de tiempo de los factores fueron estimadas utilizando mínimos cuadrados lineales y fijando el valor de λ como se describió anteriormente. Luego, $\lambda = 0.01868$.

En esta sección buscamos ajustar un Modelo de Corrección de Error Vectorial (VECM) a nuestra serie. Es decir, buscamos estimar α , β , ϕ y Γ_i para todo i tal que se satisfaga (4.6) utilizando los valores de la serie $\{X_t\}$ obtenida de Nelson y Siegel junto con las series de los drivers $\{D_t\}$. Previo a esto, necesitamos llevar a cabo un análisis de integración y cointegración para la serie de datos:

Análisis inicial

- Verificar que cada serie unidimensional sea integrada de orden 1 ó $I(1)$, tanto para los factores como para los drivers.
- Elegir el número de lags ó rezagos para incluir en el VAR.
- Verificar cointegración entre factores y drivers y definir el número de relaciones de cointegración existente.

Para llevar a cabo este análisis, se utilizarán diferentes tests estadísticos que son desarrollados en el Apéndice C. Por otro lado, este análisis se repite para diferentes combinaciones de factores y drivers y sólo se retienen aquellos en los que se concluye al menos una relación de cointegración.

A continuación describiremos los diferentes drivers a considerar en este trabajo:

- GDP: Siglas en inglés de Gross Domestic Product. Es el valor monetario de la producción de bienes y servicios de demanda final de un país o región durante un período determinado, normalmente de un año o trimestrales. En castellano se suele denominar PBI o Producto Bruto Interno.
- S&P 500: El índice S&P 500, o el índice Standard & Poor's 500, es un índice ponderado por capitalización de mercado de las 500 empresas más grandes de los Estados Unidos que cotizan en bolsa. El índice es ampliamente considerado como el mejor indicador de las acciones estadounidenses de gran capitalización.
- Slope Tsy 2/10: La pendiente de la curva de rendimiento representada a través de la diferencia entre el rendimiento del bono del Tesoro a 10 años y el rendimiento a 2 años.
- MBS: Siglas en inglés de Mortgage backed securities. Son los valores respaldados por hipotecas. El driver en sí representa la duración de dichos valores.
- USD/YEN: Tipo de cambio entre el Yen y el Dólar Estadounidense.
- Libor-Repo O/N: Diferencia entre la tasa Libor y la Tasa Repo Overnight. La tasa Libor es una tasa interbancaria (ver Sección 1.3.1) y la tasa Repo proviene de un acuerdo por el que una parte compradora cede efectivo a plazo y recibe un activo financiero (por ejemplo un bono) como garantía por parte de la contraparte, con el compromiso de que la parte vendedora lo recompre en una

fecha futura entregando el efectivo más un interés. Un pacto de recompra por un día se llama un *Overnight Repo* y la tasa de interés involucrada, *tasa Repo*.

En la siguiente Tabla 4.1 se describen algunos estadísticos de las series a analizar:

Tabla 4.1: Estadísticos de factores y drivers

Variables	Media	Desv. Estándar	Mínimo	Máximo
Nivel (factor 1)	$(-3.796) \times 10^{-1}$	2.098×10^{-1}	$(-8.567) \times 10^{-1}$	$(-2.688) \times 10^{-2}$
Pendiente (factor 2)	5.624×10^{-1}	2.554×10^{-1}	1.062×10^{-1}	1.243
Curvatura (factor 3)	8.223×10^{-1}	5.194×10^{-1}	(-1.121×10^{-1})	2.646
GDP (Driver 1)	(-6.074)	2.612	$(-1.010) \times 10^{-1}$	(-2.300)
S& P 500 (Driver 2)	1.544×10^3	3.423×10^2	1.023×10^3	2.131×10^3
Slope Tsy 2/10 (Driver 3)	2.026	1.190	2.910	4.828×10^{-1}
MBS (Driver 4)	3.611	1.153	0.900	5.600
USD/YEN (Driver 5)	9.428×10^1	1.459×10^1	7.570×10^1	1.255×10^2
Libor-Repo O/N (Driver 6)	1.720×10^{-1}	9.574×10^{-2}	$(-2.130) \times 10^{-1}$	5.820×10^{-1}

Asumimos para cada serie unidimensional un modelo autorregresivo $AR(p)$, $y_t = c + a_1y_{t-1} + a_2y_{t-2} + \dots + a_py_{t-p} + \epsilon_t$, donde el número de lags p , término constante c y los coeficientes a_k dependen del factor o driver seleccionado.

Integrabilidad: En las Figuras 4.2 a 4.7 observamos cómo las series de los factores de las curvas de Swap Spread y de los drivers a considerar parecieran no moverse en torno a una media fija o determinista.

En las primeras figuras se muestra, por otro lado, que los factores son claramente estacionarios en sus primeras diferencias; es decir, son integrados de orden 1 o $I(1)$.

Figura 4.2: Serie del factor nivel y serie de sus diferencias

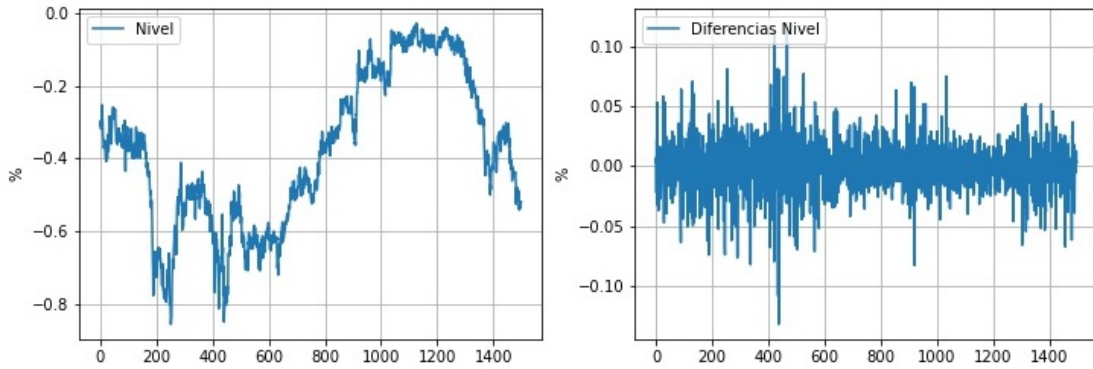


Figura 4.3: Serie del factor pendiente y serie de sus diferencias

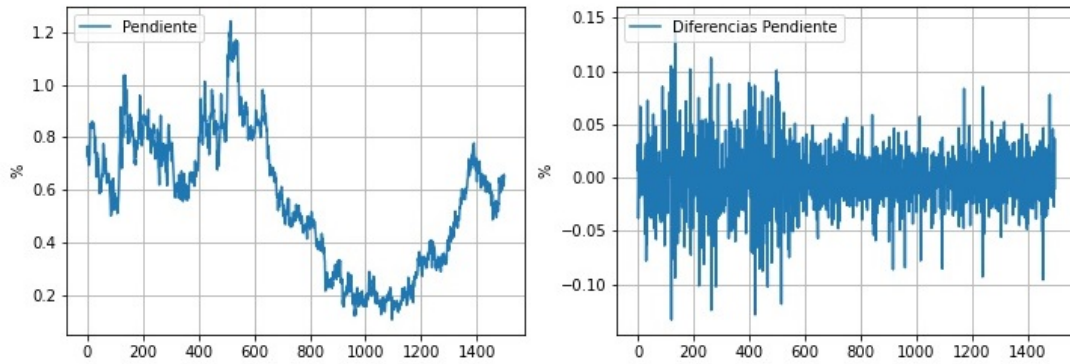
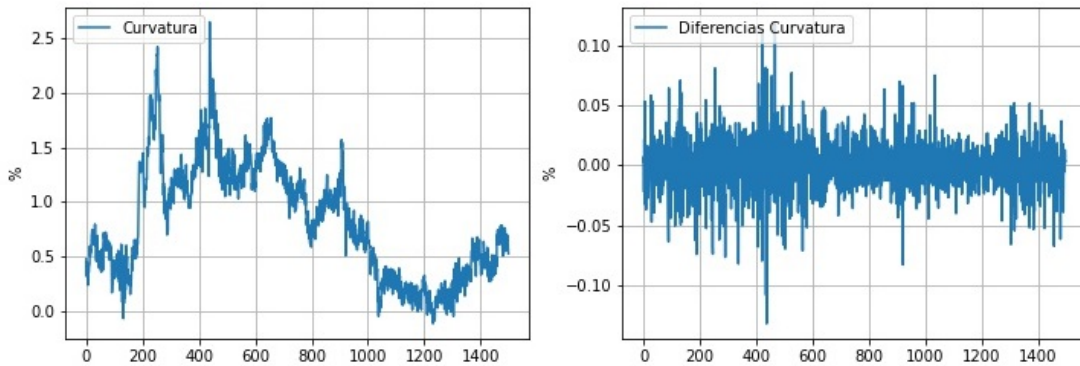


Figura 4.4: Serie del factor curvatura y serie de sus diferencias



Como vimos previamente en la Sección 3.1.2, una serie de tiempo $AR(p)$ tiene una raíz unitaria si una de las raíces del polinomio característico inverso

$$Q(x) = 1 - (a_1x + a_2x^2 + \dots + a_px^p) = (1 - \lambda_1x) \dots (1 - \lambda_px)$$

es 1. En este caso decimos que el proceso $AR(p)$ no es estable.

Si bien esto no implica que el proceso sea no estacionario (pues la recíproca no es cierta) por lo analizado en (3.1.3), vamos a suponer que ocurre una de las siguientes situaciones:

- El polinomio tiene raíces unitarias y las demás fuera del círculo unitario. En este caso el proceso no es estable y no es estacionario.
- El polinomio tiene todas las raíces fuera del círculo unitario. En este caso el proceso resulta estable, y por lo tanto es estacionario.

Luego, nos basta testear si el proceso tiene alguna raíz unitaria o no. Esto puede ser confirmado desde un punto de vista estadístico estricto en los resultados arrojados por el test Augmented Dickey Fuller (ADF), (ver Apéndice C).

La hipótesis nula del test ADF es que la variable bajo consideración tiene una raíz unitaria. La hipótesis alternativa es que la variable no tiene raíz unitaria. Los resultados del test se muestran en la Tabla 4.2.

Tabla 4.2: Test ADF de factores y drivers- Sin asumir tendencia temporal en la media.

Variables	Nivel (constante, no tendencia)	Primeras Diferencias (no constante, no tendencia)
Nivel (factor 1)	(-1.63)	(-48.79)***
Pendiente (factor 2)	(-1.96)	(-45.00)***
Curvatura (factor 3)	(-2.52)	(-35.77)***
GDP (Driver 1)	(-0.61)	(-38.69)***
S& P (Driver 2)	(-0.60)	(-39.23)***
Slope Tsy 2/10 (Driver 3)	(-1.60)	(-39.40)***
MBS (Driver 4)	(-1.80)	(-30.03)***
USD/YEN (Driver 5)	(0.35)	(-38.12)***

Nota: *, **, y *** representan significancia en los niveles al 10 %, 5 % y 1 %, respectivamente.

Para testear si una variable dada en la primer columna de la Tabla 4.2 es estacionaria, usamos el modelo que incluye una constante pero no una tendencia en el tiempo. Esto es debido a que las variables

en estudio tienen media distinta de cero. El correspondiente estadístico está dado en la segunda columna de dicha tabla.

Para testear si una variable dada en primeras diferencias es estacionaria, usamos un modelo que no incluye ni constante ni tendencia en el tiempo. Esto se debe a que las primeras diferencias de las variables tienen una media que no es significativamente diferente de cero. El correspondiente estadístico está dado en la última columna de la tabla.

La Tabla 4.2 muestra que la hipótesis de no estacionariedad no puede ser rechazada para los factores y drivers mientras que sí puede ser rechazada la hipótesis nula en las primeras diferencias en el nivel de confianza del 1 %. Entonces, podemos considerar que todas las variables están integradas de orden 1.

A continuación podemos observar los gráficos de los factores junto con los drivers que resultaron no estacionarios:

Figura 4.5: Series del nivel y drivers:

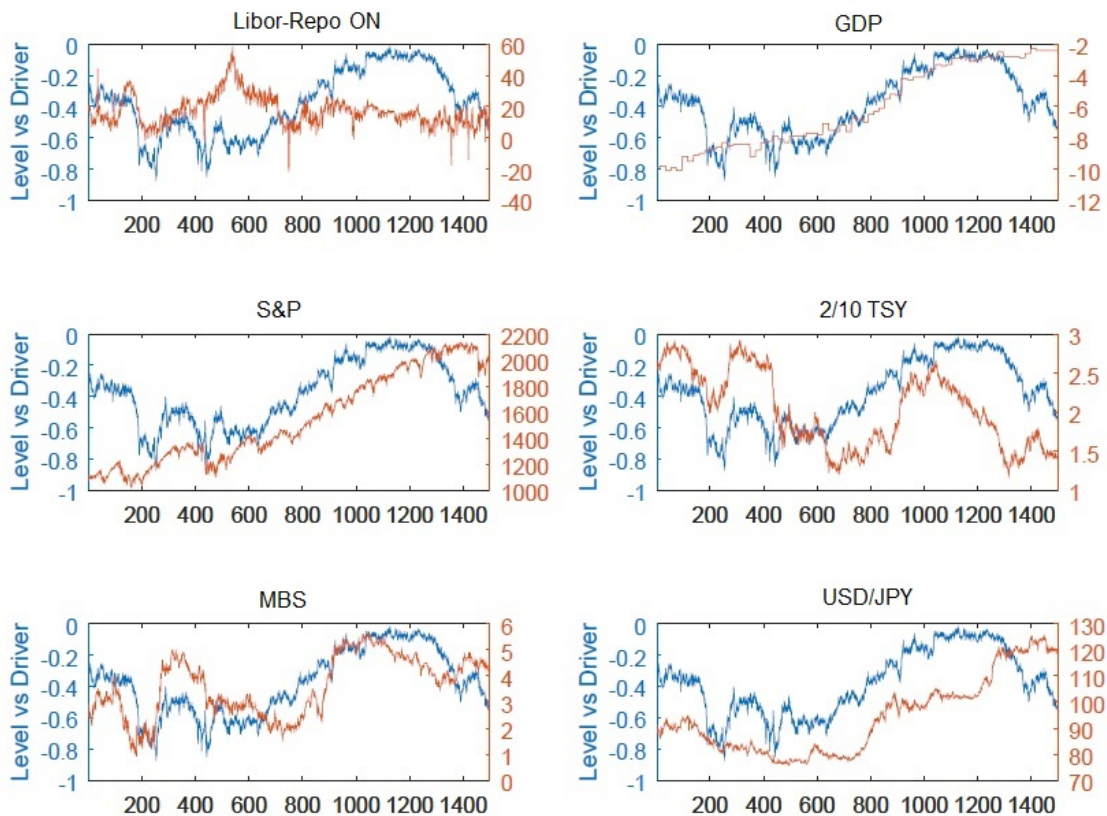


Figura 4.6: Series de la pendiente y drivers:

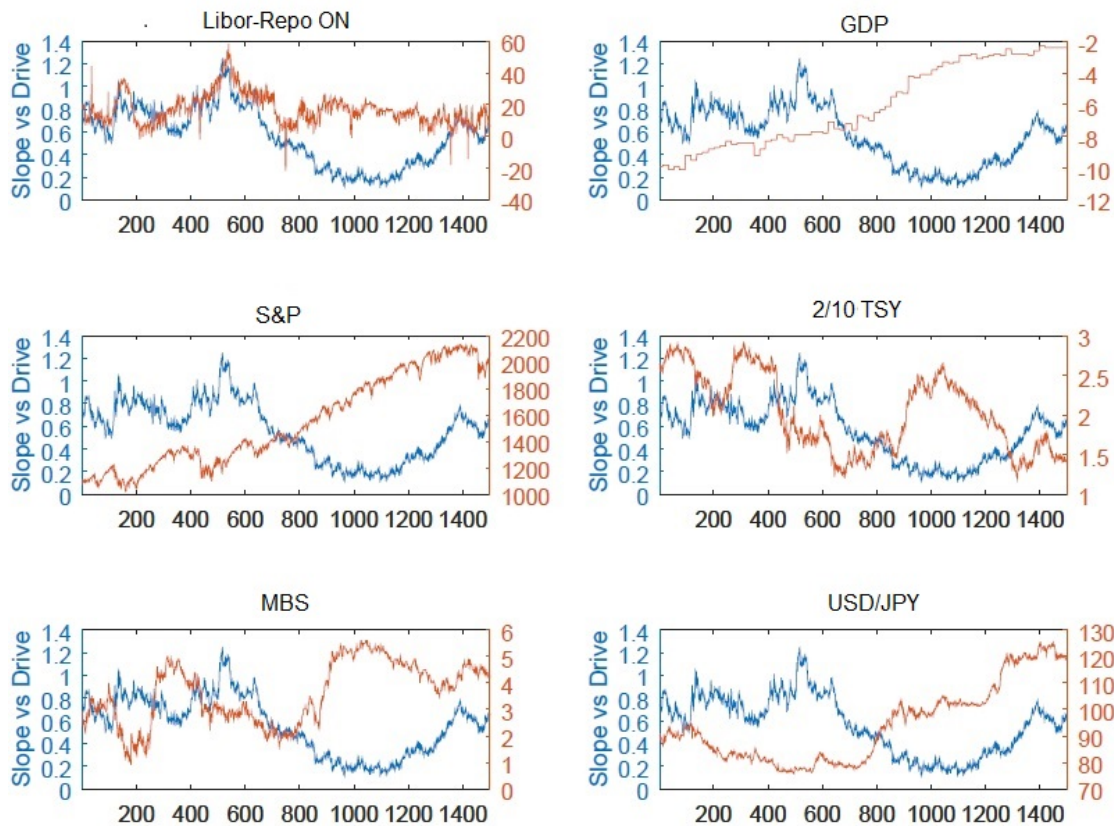
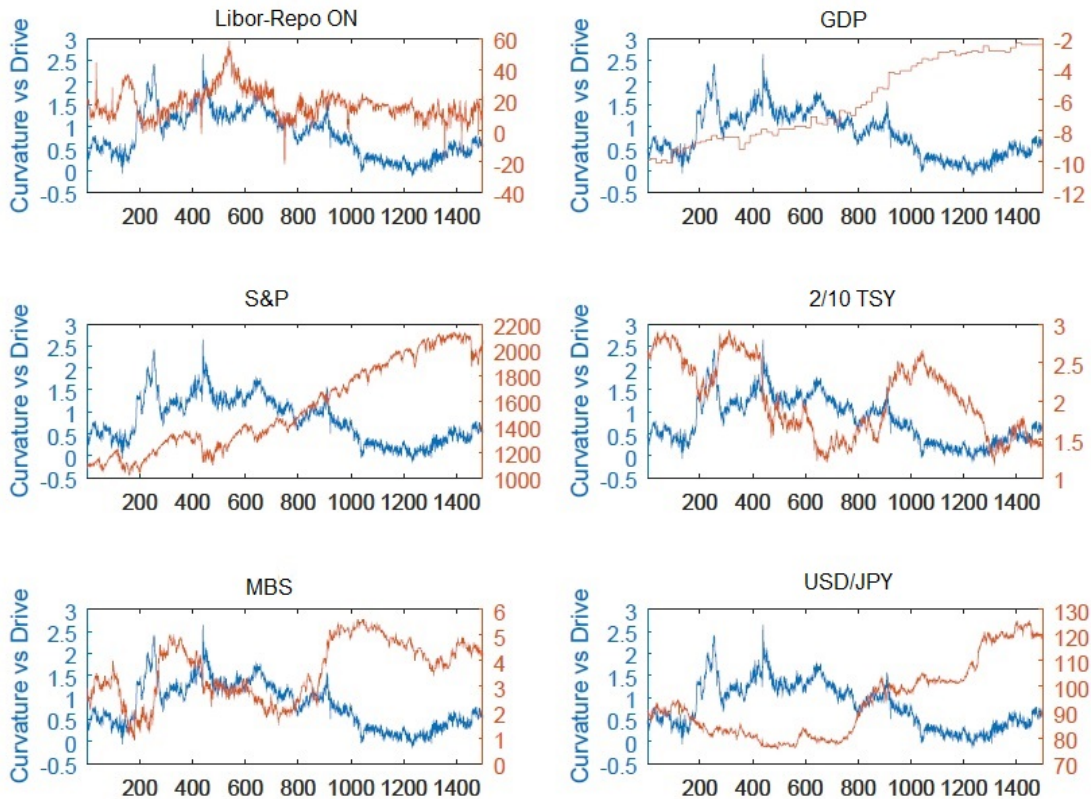


Figura 4.7: Series de la curvatura y drivers:



Especificación del retardo óptimo del VAR: Es importante definir de manera precisa la longitud del retardo pues si el retardo es muy bajo probablemente no se capture completamente la dinámica del sistema que está siendo modelado. Por otra parte, si es demasiado alto, se corre el riesgo de perder grados de libertad y tener que estimar un número muy grande de parámetros. El retardo óptimo es esencial por cuanto es la base para el cálculo del número de vectores de cointegración. Sin embargo, a los fines de conseguir cierta parsimonia en nuestro modelo, el número de retardos en el modelo VAR se determinó utilizando el criterio de información de Schwarz o Bayesian Information Criteria (BIC). Esta elección se debe principalmente a que suele ser el que arroja el lag más bajo entre todos los criterios puesto que posee un término de penalización para el número de parámetros en el modelo. Para un mayor desarrollo de este test (ver Apéndice C).

Rango de Cointegración: El objetivo en este punto es identificar al menos una relación de cointegración existente entre las variables seleccionadas. Para este propósito usamos el test de cointegración de Johansen para testear si existen relaciones a largo plazo involucrando los factores de las curvas de Swap Spread y la variables explicatorias.

El método de S. Johansen considera las siguientes pruebas para determinar el número de vectores de cointegración:

- Prueba de la Traza:

La hipótesis nula para este test es que el número de vectores de cointegración es $r = r^* < k$, frente a la alternativa de que $r = k$. La prueba se realiza de manera sucesiva para $r^* = 0, 1, 2$, etc., tal que el estimador para r se obtiene del primer no rechazo de la hipótesis nula.

- Prueba del Máximo Valor Propio:

La alternativa es $r = r^* + 1$ y, nuevamente, la prueba se realiza de forma sucesiva para $r^* = 0, 1, 2$, etc., tal que el estimador para r se obtiene del primer no rechazo.

En el Apéndice C se describen más detalles de ambos tests.

En la siguiente sección abordaremos 3 combinaciones de factores y drivers donde se puede ver una clara cointegración entre ellos. Allí mostraremos los resultados del test de la traza para cada caso.

4.5.3. Implementación en Python

Todo el trabajo se implementó en lenguaje Python y se utilizó el paquete *statsmodels.tsa.vector_ar* desarrollado por Seabold and Perktold [54]. La documentación se puede encontrar aquí: <https://www.statsmodels.org/stable/>.

El código para obtener los resultados del trabajo se encuentra subido al repositorio *Swap-Spread-Tesis* de Git Hub. El link al mismo es el siguiente: <https://github.com/Elisa-Ravasi/Swap-Spread-Tesis>.

Modelo DNS-VAR

Para la estimación de parámetros de nuestro modelo VAR se utilizaron las funciones del paquete https://www.statsmodels.org/stable/_modules/statsmodels/tsa/vector_ar/var_model.html. Específicamente, utilizamos las funciones `model.select_order()` para obtener el lag p , `model.fit()` para ajustar los parámetros del modelo y `model.forecast()` para predicción, donde *model* es una clase VAR. La clase VAR asume que las series de tiempo son estacionarias. La función `model.fit()` devuelve los parámetros del modelo obtenidos de las estimaciones realizadas por el método de cuadrados mínimos.

Modelo DNS-VECM

Como en el modelo anterior, se utilizó la función `model.select_order()` para obtener el lag p del modelo VAR y la función `adfuller()` del paquete `statsmodels.tsa.stattools` para analizar raíces unitarias en factores y drivers.

Para la estimación de parámetros de nuestro modelo VECM, se utilizaron las funciones del paquete https://www.statsmodels.org/stable/_modules/statsmodels/tsa/vector_ar/vecm.html implementado por J. Perktold, S. Seabold y J. Taylor. Específicamente, utilizamos la función `model.fit()` donde `model` es una clase VECM. Esta función devuelve los parámetros del modelo obtenidos de las estimaciones realizadas por el método de máxima verosimilitud.

Para poder obtener el rango de cointegración de cada conjunto de variables analizadas en nuestro trabajo, utilizamos la función `coint_johansen(y, det_order, k_ar_diff)`. Esta función implementa ambos tests de Johansen y devuelve el estadístico para cada test junto con el nivel de significación.

Como se puede observar en la siguiente discusión: <https://github.com/statsmodels/statsmodels/issues/5355>, dicha función tiene un bug que se presenta a la hora de definir el input `det_order`.

Nosotros reportamos este bug que luego fue corregido. Esto se puede observar en el siguiente link: https://www.statsmodels.org/dev/_modules/statsmodels/tsa/vector_ar/vecm.html#coint_johansen. El ticker GH 5731 identifica nuestra observación.

```

641 642
642 643     dx = detrend(dx, f)
643 644     r0t = resid(dx, z)
644 -     lx = endog[:-k_ar_diff]
645 +     # GH 5731, [:-0] does not work, need [:t-0]
646 +     lx = endog[:endog.shape[0]-k_ar_diff]
647 647     lx = lx[1:]
648 648     dx = detrend(lx, f)
649 649     rkt = resid(dx, z) # level on lagged diffs
650 +     # Level covariance after filtering k_ar_diff
651 651     skk = np.dot(rkt.T, rkt) / rkt.shape[0]
652 +     # Covariance between filtered and unfiltered
653 653     sk0 = np.dot(rkt.T, r0t) / rkt.shape[0]
654 654     s00 = np.dot(r0t.T, r0t) / r0t.shape[0]
655 655     sig = np.dot(sk0, np.dot(inv(s00), sk0.T))

```

Para definir nuestra clase VECM debimos pasar el argumento `deterministic` que indica si se desea asumir términos deterministas en el modelo. Para los dos casos especiales de un término determinista dentro de la relación de cointegración: un intercepto y una tendencia lineal, podemos pasar “ci” y “li”, respectivamente al argumento `deterministic`. También podemos usar términos deterministas fuera

de la relación de cointegración. Especificamos dichos términos al pasar “co” para asumir intercepto y “lo” para tendencia lineal.

La siguiente Tabla (4.3) muestra los cinco casos considerados en Johansen [35]. La última columna indica qué cadena pasar al argumento `deterministic` para cada uno de estos casos:

Tabla 4.3: Casos considerados para el test de Johansen, Johansen [35]

Caso	Intercepto	Pendiente de la tendencia lineal	deterministic
I	0	0	“nc”
II	$-\alpha\beta^T\mu$	0	“ci”
III	$\neq 0$	0	“co”
IV	$\neq 0$	$-\alpha\beta^T\gamma$	“coli”
V	$\neq 0$	$\neq 0$	“colo”

Por ultimo, realizamos un estudio de simulación que detallamos en la Sección 4.6

4.6. Simulación y resultados

En esta sección detallaremos los resultados obtenidos con ambos modelos VAR y VECM para factores y Drivers.

Llevaremos a cabo 3 estudios de simulación:

Caso A: Consideraremos un VAR formado por el primer factor y el driver MBS.

Caso B: Consideraremos un VAR formado por el segundo factor y el driver USD/JPY.

Caso C: Consideraremos un VAR formado por el tercer factor y el driver MBS.

Es importante mencionar que los valores del driver MBS corresponden al MBS effective duration. “Effective duration” expresa el cambio medible en el valor de un derivado que depende de tasas de interés en respuesta a un cambio en dichas tasas.

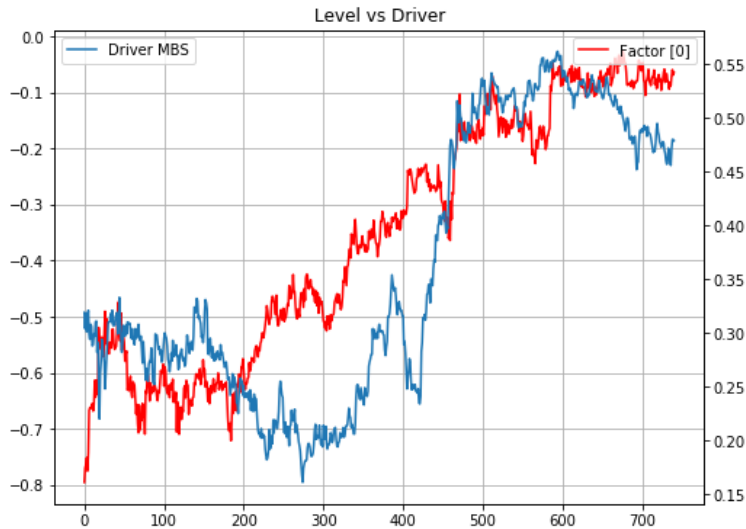
Ejemplo: Si un bono tiene una effective duration de 8.75, significa que si hubiera un cambio en el rendimiento del bono de un 1 %, se esperaría que el precio del bono cambie en 8.75 %.

4.6.1. Caso A: Nivel y MBS

Para este primer caso, se consideraron los datos del factor 1 (nivel) y los datos del driver MBS del 25/08/2011 al 04/08/2014, que corresponde a una muestra de tamaño 741. Esto se debe a la notoria

relación de equilibrio a largo plazo entre ambas series, como se puede observar en la Figura 4.8. Para evitar problemas con las unidades, la serie MBS se multiplicó por 0.1.

Figura 4.8: *Factor 1 y driver MBS* - Período: 25/08/2011 a 04/08/2014.



Para ajustar a un modelo VAR la serie multivariada

$$\tilde{X}_t = \begin{pmatrix} L_t \\ \text{MBS}_t \end{pmatrix},$$

estimamos los parámetros correspondientes a partir de los primeros 740 datos. El número estimado de rezagos fue $p = 2$ y los parámetros del modelo VAR(2) obtenidos de la muestra fueron los siguientes:

$$A1 = \begin{bmatrix} 0.679 & 0.060 \\ (-0.029) & 0.991 \end{bmatrix} \quad A2 = \begin{bmatrix} 0.311 & (-0.049) \\ 0.036 & (-0.004) \end{bmatrix}$$

$$\Sigma_u = \begin{bmatrix} 3.738 \times 10^{-4} & (-4.616) \times 10^{-6} \\ (-4.616) \times 10^{-6} & 8.805 \times 10^{-5} \end{bmatrix} \quad C = \begin{bmatrix} (-0.007) \\ 0.008 \end{bmatrix}$$

Con estos parámetros, ejecutamos un estudio de simulación para predecir el valor del nivel L_{741} con una tolerancia (tol) de 0.03, 0.01 y 0.003, respectivamente. La Tabla 4.4 resume los resultados de la simulación.

Nota. La media estimada corresponde a la media muestral obtenida con los datos de predicción para el valor en T en cada simulación. La std muestral corresponde a la desviación estándar de ese estimador.

En la Figura 4.9 podemos ver la trayectoria del proceso del factor 1 (asumiendo un VAR) durante la última simulación para $tol = 0.03$, $tol = 0.01$ y $tol = 0.003$.

Tabla 4.4: Resultados del modelo VAR que incluye el factor 1 o nivel y driver MBS con datos diarios desde el 25/08/2011 hasta el 04/08/2014.

Resultados de Predicción L_{741}	$tol = 0.03$	$tol = 0.01$	$tol = 0.003$
Lag VAR (BIC)	2	2	2
Número de Simulaciones	100	327	3440
Media Estimada Factor 1 (%)	(-0.089)	(-0.118)	(-0.123)
Sesgo Factor 1	(-0.025)	(-0.054)	(-0.058)
Media Estimada Driver MBS (Dur)	0.519	0.496	0.495
Sesgo Driver MBS	0.040	0.017	0.016
Std Muestral Factor 1	0.019	0.01	0.003
Std Muestral Driver MBS	0.011	0.006	0.002
Tiempo Ejecución	21 s.	112 s.	937 s.

Para la estimación del modelo VECM para esta misma muestra de datos, asumimos un intercepto y una tendencia lineal fuera de la ecuación de cointegración (`deterministic = 'colo'`). Las simulaciones se realizaron a partir de un modelo de cointegración para el cual se estimaron los parámetros asumiendo un rango de cointegración igual a 1. Este rango se determinó mediante el test de traza con un nivel de confianza del 95 % y el parámetro `det_order` se tomó igual a 1 asumiendo un término de tendencia.

Los parámetros estimados del modelo VECM para la muestra fueron los siguientes:

$$\alpha = \begin{bmatrix} (-0.043) \\ 0.017 \end{bmatrix} \quad \beta = \begin{bmatrix} 1 \\ (-0.501) \end{bmatrix}$$

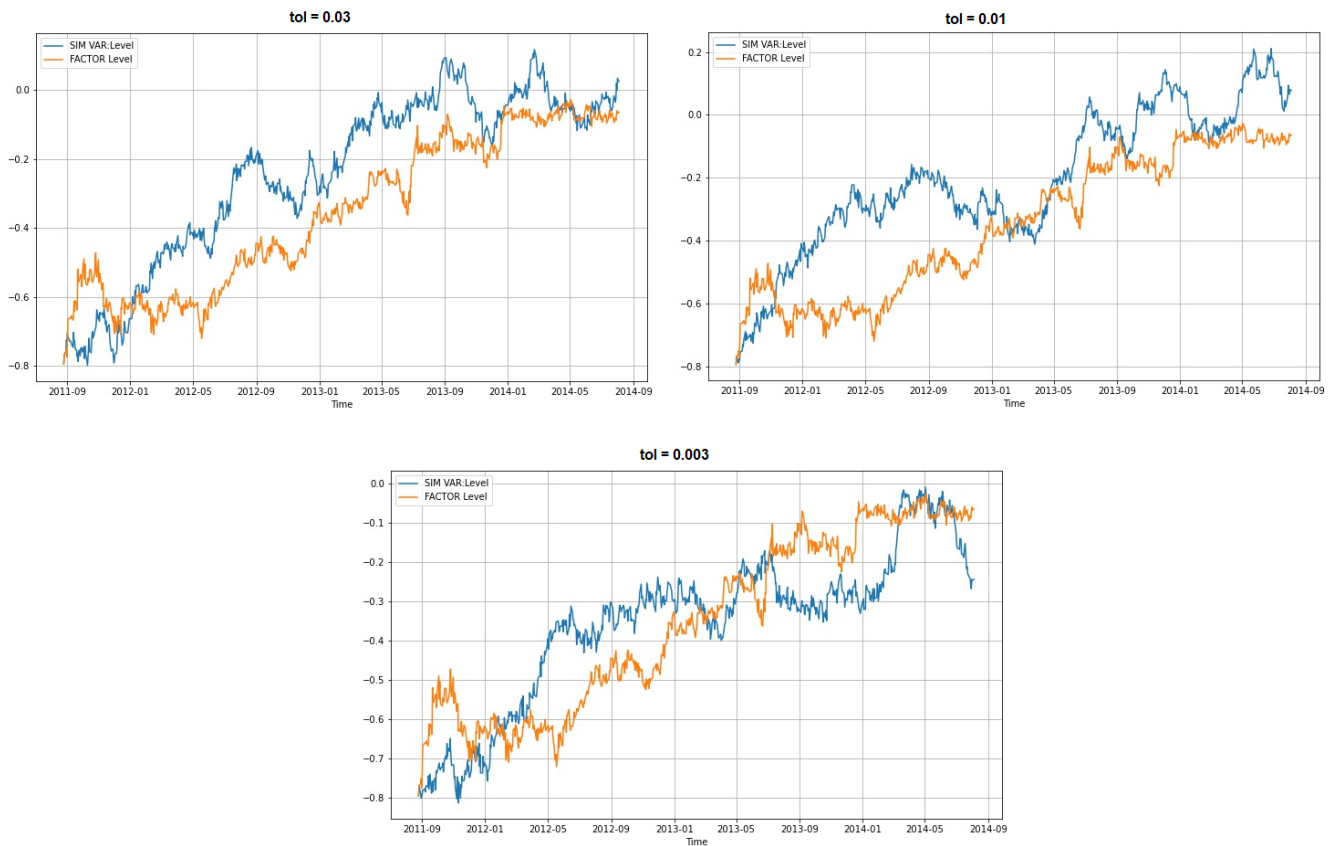
$$\Gamma_1 = \begin{bmatrix} (-0.294) & 0.052 \\ (-0.041) & 0.001 \end{bmatrix} \quad C = \begin{bmatrix} (-3.301) \times 10^{-2} & 3.029 \times 10^{-5} \\ 1.371 \times 10^{-2} & (-1.197) \times 10^{-5} \end{bmatrix}$$

$$\Sigma_u = \begin{bmatrix} 3.671 \times 10^{-4} & (-3.257) \times 10^{-6} \\ (-3.257) \times 10^{-6} & 8.743 \times 10^{-5} \end{bmatrix}$$

La Tabla 4.5 resume los valores obtenidos.

Nota. La media estimada corresponde a la media muestral obtenida con los datos de predicción para el valor en T en cada simulación. La std muestral corresponde a la desviación estándar de ese estimador.

Figura 4.9: Trayectoria del Factor 1-Caso A (VAR): Muestra Real (Factor Nivel) y última simulación (Nivel VAR).



Observación. Podemos observar que el modelo VAR necesitó entre un 60 % y un 70 % más de simulaciones que el modelo VECM para alcanzar la tolerancia pretendida en los casos $tol = 0.01$ y $tol = 0.003$. Y entre un 20 % y un 40 % más en tiempo de ejecución.

El test ADF asumiendo una regresión con constante y tendencia (ct) para el proceso $\beta \tilde{X}_t$, arroja los resultados descritos en La Tabla 4.6. Recordar que \tilde{X}_t es el proceso formado por el factor 1 y el driver MBS. Es decir, el estadístico del test ADF (ct) para el proceso $\beta \tilde{X}_t$ es (-4.014) con p -valor igual a 0.008. Entonces, podemos rechazar la hipótesis de raíz unitaria para este proceso estacionario con tendencia con un nivel de significancia del 1 %.

En la Figura 4.10 podemos observar la trayectoria de la serie de tiempo del factor 1 (considerando un

Tabla 4.5: Resultados del modelo VECM que incluye el factor 1 o nivel y el driver MBS con datos diarios desde el 25/08/2011 hasta el 04/08/2014 asumiendo intercepto y tendencia fuera de la ecuación de cointegración.

Resultados de Predicción L_{741}	$tol = 0.03$	$tol = 0.01$	$tol = 0.003$
Lag VECM (BIC)	2	2	2
Número of Simulaciones	100	189	2128
Media Estimada Factor 1 (%)	(-0.029)	(-0.025)	(-0.037)
Sesgo Factor 1	0.036	0.040	0.027
Media Estimada Driver MBS (Dur)	0.486	0.494	0.465
Sesgo Driver MBS	0.007	0.015	(-0.014)
Std Muestral Factor 1	0.013	0.008	0.002
Std Muestral Driver MBS	0.023	0.019	0.006
Tiempo Ejecución	22 s.	91 s.	670 s.

Tabla 4.6: Test ADF (ct) para VECM - Caso A

Estadístico del test	(-4.014)		
p -valor	0.008		
Valores críticos del estadístico en los niveles	1 % (-3.971)	5 % (-3.417)	10 % (-3.131)

VECM) durante la última simulación para $tol = 0.03$, $tol = 0.01$ y $tol = 0.003$.

Finalmente, la Figura 4.11 muestra el comportamiento de la serie $\beta\tilde{X}_t$ junto con la trayectoria de la tendencia obtenida por cuadrados mínimos para esa serie.

El test ADF sin asumir término de constante ni tendencia (nc) para la serie $\{\beta\tilde{X}_t - ((-0.812) + 0.001t)\}_t$ se describe en La Tabla 4.7.

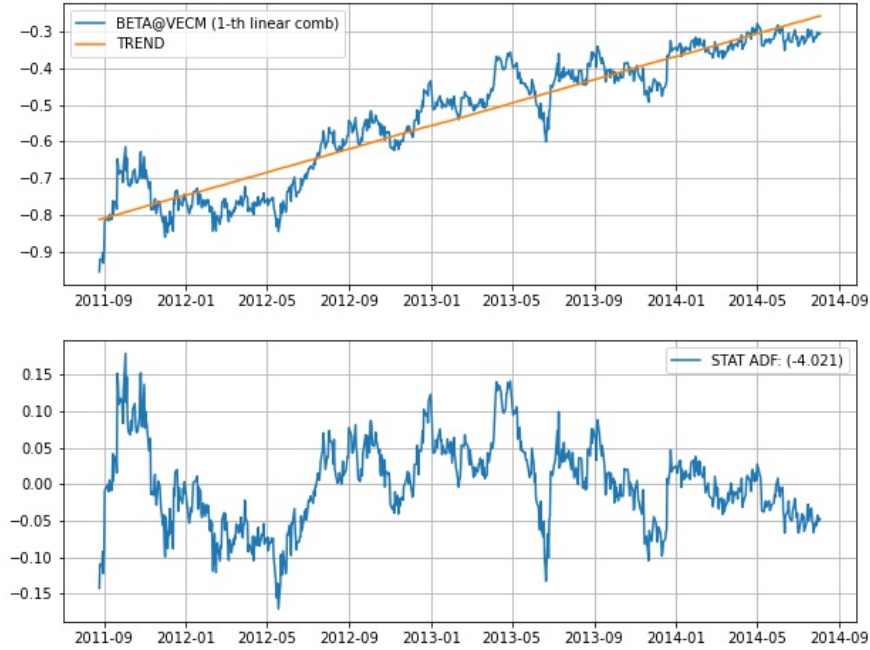
Tabla 4.7: Test ADF (nc) - Caso A

Estadístico del test	(-4.021)		
p -valor	1.000×10^{-4}		
Valores críticos del test a los niveles	1 % (-2.569)	5 % (-1.941)	10 % (-1.617)

Es decir, el estadístico del test viene dado por (-4.021), lo que nos permite afirmar que la hipótesis de raíz unitaria se rechaza para ese proceso con un nivel de significancia del 1 %.

Figura 4.10: *Trayectoria del Factor 1-Caso A (VECM): Muestra Real (Factor Nivel) y última simulación (Nivel VECM).*



Figura 4.11: Trayectoria $\beta\tilde{X}_t$: Tendencia (Factor Nivel-MBS) y su diferencia.

4.6.2. Caso B: Pendiente y USD/JPY

Para este segundo análisis se consideraron los datos del factor 2 y los datos del driver USD/JPY del 28/03/2013 al 13/08/2015, obteniendo una muestra de tamaño 600. Esto se debe a la notoria relación de equilibrio a largo plazo entre ambas series, como se puede observar en la Figura 4.12. Además, para evitar problemas con las unidades, multiplicaremos por 0.01 la serie USD/JPY.

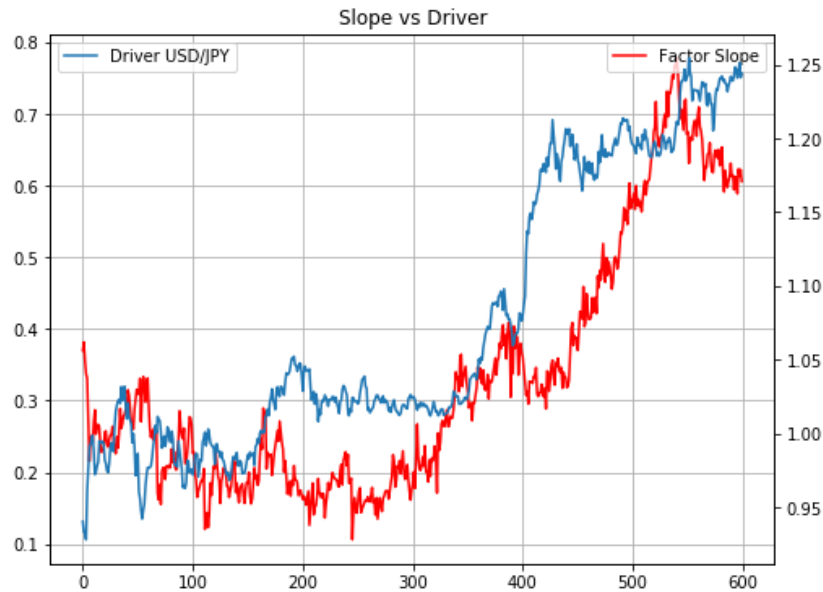
Para la determinación del modelo VAR correspondiente a la serie multivariada

$$\tilde{X}_t = \begin{pmatrix} S_t \\ \text{JPY}_t \end{pmatrix},$$

se estimaron los parámetros utilizando los primeros 599 valores de la muestra. El número estimado de rezagos fue $p = 2$ y los parámetros del modelo VAR(2) obtenidos desde la muestra fueron los siguientes:

$$A_1 = \begin{bmatrix} 0.708 & (-0.218) \\ 0.004 & 1.027 \end{bmatrix} \quad A_2 = \begin{bmatrix} 0.260 & 0.282 \\ 0.002 & (-0.038) \end{bmatrix}$$

$$\Sigma_u = \begin{bmatrix} 4.441 \times 10^{-4} & (-7.616) \times 10^{-6} \\ (-7.616) \times 10^{-6} & 3.821 \times 10^{-5} \end{bmatrix} \quad C = \begin{bmatrix} (-0.057) \\ 0.011 \end{bmatrix}.$$

Figura 4.12: *Factor 2 and driver USD/JPY* - Período: 28/03/2013 to 13/08/2015.

Luego, se realizaron simulaciones para la predicción de la pendiente S_{600} para obtener una desviación del orden (tol) de 0.03, 0.01 y 0.003, respectivamente. La Tabla 4.8 resume los valores obtenidos.

Nota. La media estimada corresponde a la media muestral obtenida con los datos de predicción para el valor en T en cada simulación. La std muestral corresponde a la desviación estándar de ese estimador.

La Figura 4.13 muestra la trayectoria del proceso del factor 2 durante la última simulación para $tol = 0.03$, $tol = 0.01$ y $tol = 0.003$.

Para la estimación del modelo VECM para esta misma muestra de datos, asumimos un intercepto y una tendencia lineal fuera de la ecuación de cointegración (`deterministic = 'colo'`). Las simulaciones se realizaron a partir de un modelo de cointegración para el cual se estimaron los parámetros asumiendo un rango de cointegración igual a 1. Este rango se determinó mediante la prueba de trazas con un nivel de confianza del 95 % y el parámetro `det_order` se tomó igual a 1 asumiendo un término de tendencia.

Los parámetros estimados del modelo VECM para la muestra fueron los siguientes:

$$\alpha = \begin{bmatrix} (-0.030) \\ 0.007 \end{bmatrix} \quad \beta = \begin{bmatrix} 1 \\ (-2.137) \end{bmatrix}$$

Tabla 4.8: Resultados del VAR que incluyen el factor 2 o pendiente y el driver USD/JPY con data diaria desde 28/03/2013 to 13/08/2015.

Resultados de Predicción S_{600}	$tol = 0.03$	$tol = 0.01$	$tol = 0.003$
Lag VAR (BIC)	2	2	2
Número de Simulaciones	100	699	7468
Media Estimada Factor 2 (%)	0.615	0.648	0.644
Sesgo Factor 2	0.010	0.042	0.038
Media Estimada Driver USD/JPY	1.236	1.250	1.248
Sesgo Driver USD/JPY	(-0.008)	0.006	0.004
Std Muestral Factor 2	0.028	0.01	0.003
Std Muestral Driver USD/JPY	0.015	0.005	0.002
Tiempo Ejecución	25 s.	111 s.	1201 s.

$$\Gamma_1 = \begin{bmatrix} (-0.261) & (-0.284) \\ (-0.003) & 0.039 \end{bmatrix} \quad C = \begin{bmatrix} (-5.812) \times 10^{-2} & (-4.112) \times 10^{-7} \\ 1.329 \times 10^{-2} & 2.310 \times 10^{-6} \end{bmatrix}$$

$$\Sigma_u = \begin{bmatrix} 4.407 \times 10^{-4} & (-7.590) \times 10^{-6} \\ (-7.590) \times 10^{-6} & 3.779 \times 10^{-5} \end{bmatrix}$$

Los valores obtenidos se resumen en la Tabla 4.9.

Nota. La media estimada corresponde a la media muestral obtenida con los datos de predicción para el valor en T en cada simulación. La std muestral corresponde a la desviación estándar de ese estimador.

La Figura 4.14 muestra la trayectoria del proceso del factor 2 durante la última simulación para $tol = 0.03$, $tol = 0.01$ y $tol = 0.003$.

Figura 4.13: Trayectoria del Factor 2-Caso B (VAR): Muestra Real (Factor Pendiente) y última simulación (VAR Pendiente).



Observación. Podemos observar que el modelo VAR necesitó un 30 % y un 10 % más de simulaciones para alcanzar la tolerancia pretendida en los casos $tol = 0.01$ y $tol = 0.003$, respectivamente. El tiempo de ejecución para ambos modelos fue similar.

El test ADF asumiendo una regresión con constante y tendencia (ct) para el proceso $\beta \tilde{X}_t$, arroja los resultados descriptos en La Tabla 4.10. Recordar que \tilde{X}_t es el proceso formado por el factor 2 y el driver USD/JPY.

Es decir, el estadístico del test ADF (ct) para el proceso $\beta \tilde{X}_t$ es (-3.842) con p -valor igual a 0.015. Entonces, podemos rechazar la hipótesis de raíz unitaria para este proceso estacionario con tendencia con un nivel de significancia del 5 %.

Tabla 4.9: Resultados de VECM que incluyen el factor 2 o pendiente y el driver USD/ JPY con datos diarios desde el 28/03/2013 hasta el 13/08/2015 asumiendo intercepto y tendencia fuera de la ecuación de cointegración.

Resultados Predicción S_{600}	$tol = 0.03$	$tol = 0.01$	$tol = 0.003$
Lag VECM (BIC)	2	2	2
Número de Simulaciones	100	523	6763
Media Estimada Factor 2 (%)	0.669	0.608	0.623
Sesgo Factor 2	0.064	0.003	0.017
Media Estimada Driver USD/JPY	1.263	1.234	1.240
Sesgo Driver USD/JPY	0.019	(-0.010)	(-0.004)
Std Muestral Factor 2	0.026	0.009	0.003
Std Muestral Driver USD/JPY	0.012	0.005	0.001
Tiempo Ejecución	24 s.	91 s.	1216 s.

Tabla 4.10: Test ADF (ct) para VECM - Caso B

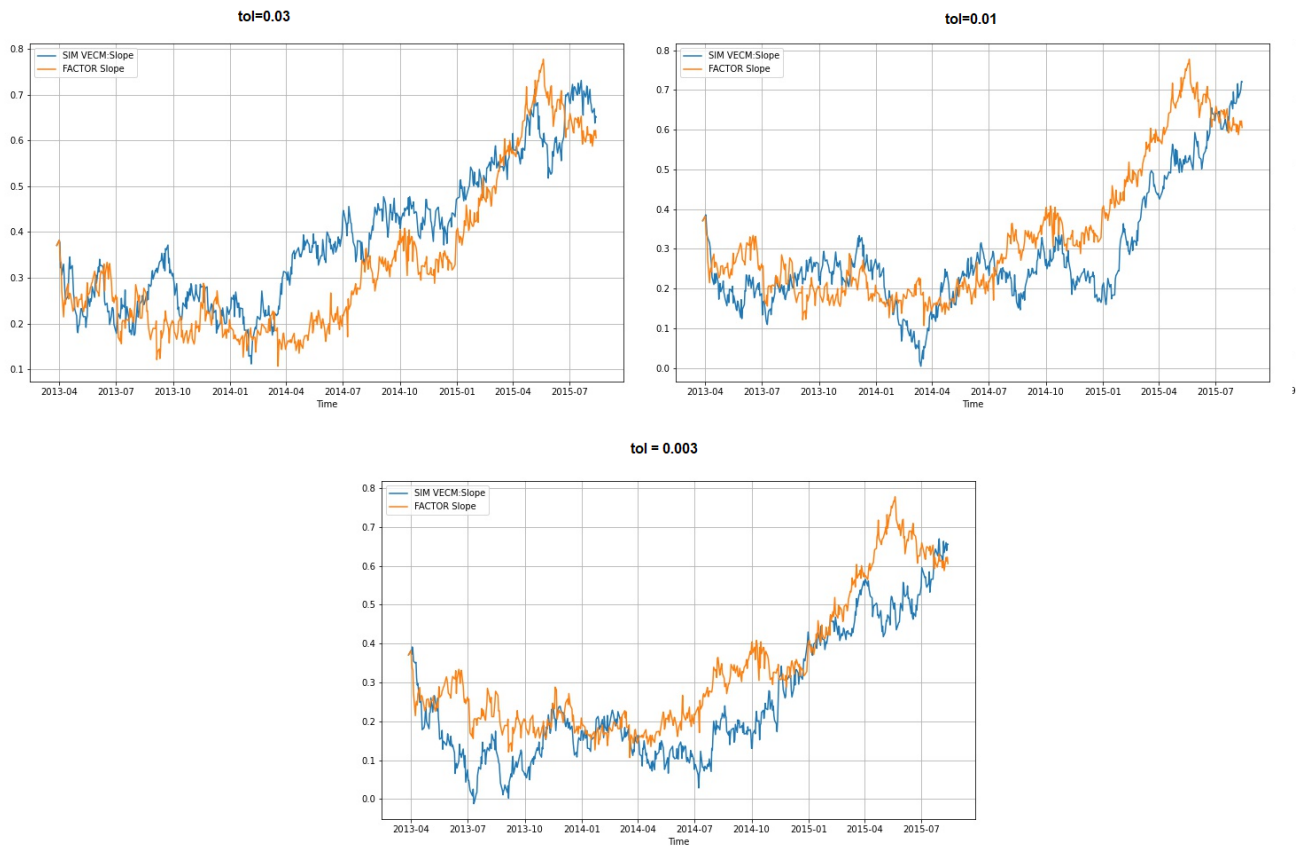
Estadístico test	(-3.842)		
p -valor	0.015		
Valores críticos del estadístico en los niveles	1 % (-3.974)	5 % (-3.418)	10 % (-3.131)

Finalmente, en la Figura 4.15 podemos observar el comportamiento de la serie $\beta\tilde{X}_t$ junto con la trayectoria de la tendencia obtenida por cuadrados mínimos para esa serie.

El test ADF sin asumir término de constante ni tendencia (nc) para la serie $\{\beta\tilde{X}_t - ((-1.887) - 2.700 \times 10^{-4}t)\}_t$ se describe en La Tabla 4.11.

Es decir, el estadístico del test viene dado por (-3.843), lo que nos permite afirmar que la hipótesis de raíz unitaria se rechaza para ese proceso con un nivel de significancia del 1 %.

Figura 4.14: *Trayectoria del Factor 2-Case B (VECM): Muestra Real (Factor Pendiente) y última simulación (VECM Pendiente).*



4.6.3. Caso C: Curvatura y MBS

Para este tercer análisis se consideraron los datos del factor 3 y los datos del driver MBS del 12/06/2012 al 15/08/2014, obteniendo una muestra de tamaño 550. Esto se debe a la notoria relación de equilibrio a largo plazo entre ambas series, como se puede observar en la Figura 4.16. Además, para evitar problemas con las unidades, multiplicaremos por 0.1 la serie MBS.

Para la determinación del modelo VAR para la serie multivariada

$$\tilde{X}_t = \begin{pmatrix} C_t \\ \text{MBS}_t \end{pmatrix},$$

se realizó la estimación de parámetros a partir de los primeros 549 valores. El número estimado de rezagos fue $p = 2$ y los parámetros del modelo VAR(2) obtenidos de la muestra fueron los siguientes:

Tabla 4.11: Test ADF (nc) - Caso B

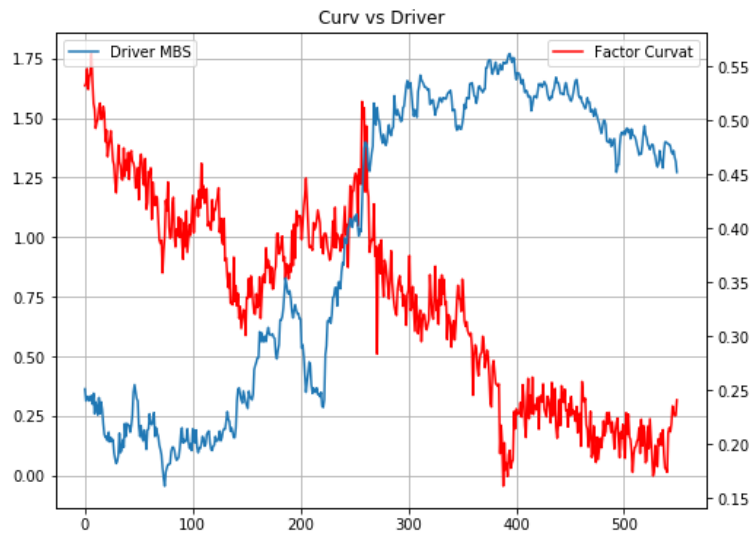
Estadístico del Test	(-3.843)		
p -valor	1×10^{-4}		
Valores críticos para el estadístico en los niveles	1 % (-2.570)	5 % (-1.942)	10 % (-1.616)

Figura 4.15: Trayectoria de $\beta\tilde{X}_t$: Tendencia (Factor Pendiente-JPY) y su serie de diferencias.

$$A1 = \begin{bmatrix} 0.531 & (-0.969) \\ 0.009 & 1.051 \end{bmatrix} \quad A2 = \begin{bmatrix} 0.426 & 0.869 \\ -0.008 & -0.051 \end{bmatrix}$$

$$\Sigma_u = \begin{bmatrix} 7.367 \times 10^{-3} & -4.350 \times 10^{-6} \\ -4.350 \times 10^{-6} & 6.861 \times 10^{-5} \end{bmatrix} \quad C = \begin{bmatrix} 0.067 \\ (-3.000) \times 10^{-4} \end{bmatrix}.$$

Figura 4.16: Factor 3 y driver MBS - Período: 12/06/2012 al 15/08/2014.



Luego, se realizaron simulaciones para la predicción del valor de la curvatura C_{550} para obtener una desviación de orden (tol) de 0.03, 0.01 y 0.003, respectivamente. La Tabla 4.12 resume los valores obtenidos.

Nota. La media estimada corresponde a la media muestral obtenida con los datos de predicción para el valor en T en cada simulación. La std muestral corresponde a la desviación estándar de ese estimador.

Tabla 4.12: Resultados del VAR que incluye el factor 3 o curvatura y el driver MBS con datos diarios desde 12/06/2012 a 15/08/2014.

Resultados de Predicción C_{550}	$tol = 0.03$	$tol = 0.01$	$tol = 0.003$
Lag VAR (BIC)	2	2	2
Número de Simulaciones	131	1222	13413
Media Estimada Factor 3 (%)	0.494	0.501	0.498
Sesgo Factor 3	0.183	0.190	0.186
Media Estimada Driver MBS (Dur)	0.459	0.458	0.460
Sesgo Driver MBS	0.007	0.006	0.008
Std Muestral Factor 3	0.030	0.01	0.003
Std Muestral Driver MBS	0.011	0.003	0.001
Tiempo Ejecución	19 s.	170 s.	1827 s

En las en los gráficos de la Figura 4.17 podemos ver la trayectoria del proceso del factor 3 durante la última simulación para $tol = 0.03$, $tol = 0.01$ y $tol = 0.003$.

Figura 4.17: *Trayectoria del Factor 3-Case C (VAR): Muestra Real (Factor Curvatura) y última simulación (VAR Curvatura).*



Para la estimación del modelo VECM para esta misma muestra de datos, asumimos un intercepto y una tendencia lineal fuera de la ecuación de cointegración (deterministic = 'colo'). Las simulaciones se realizaron a partir de un modelo de cointegración para el cual se estimaron los parámetros asumiendo un rango de cointegración igual a 1. Este rango se determinó mediante el test de la traza con un nivel de confianza del 95 % y el parámetro `det_order` se tomó igual a 1 asumiendo un término de tendencia.

Los parámetros estimados del modelo VECM para la muestra fueron los siguientes:

$$\alpha = \begin{bmatrix} (-0.087) \\ 0.004 \end{bmatrix} \quad \beta = \begin{bmatrix} 1 & (-0.308) \end{bmatrix}$$

$$\Gamma_1 = \begin{bmatrix} (-0.401) & (-0.880) \\ 0.006 & 0.047 \end{bmatrix} \quad C = \begin{bmatrix} 1.104 \times 10^{-1} & (-2.198) \times 10^{-4} \\ (-4.504) \times 10^{-3} & 9.003 \times 10^{-6} \end{bmatrix}$$

$$\Sigma_u = \begin{bmatrix} 7.222 \times 10^{-3} & 2.405 \times 10^{-6} \\ 2.405 \times 10^{-6} & 6.771 \times 10^{-5} \end{bmatrix}$$

La Tabla 4.13 resume los valores obtenidos.

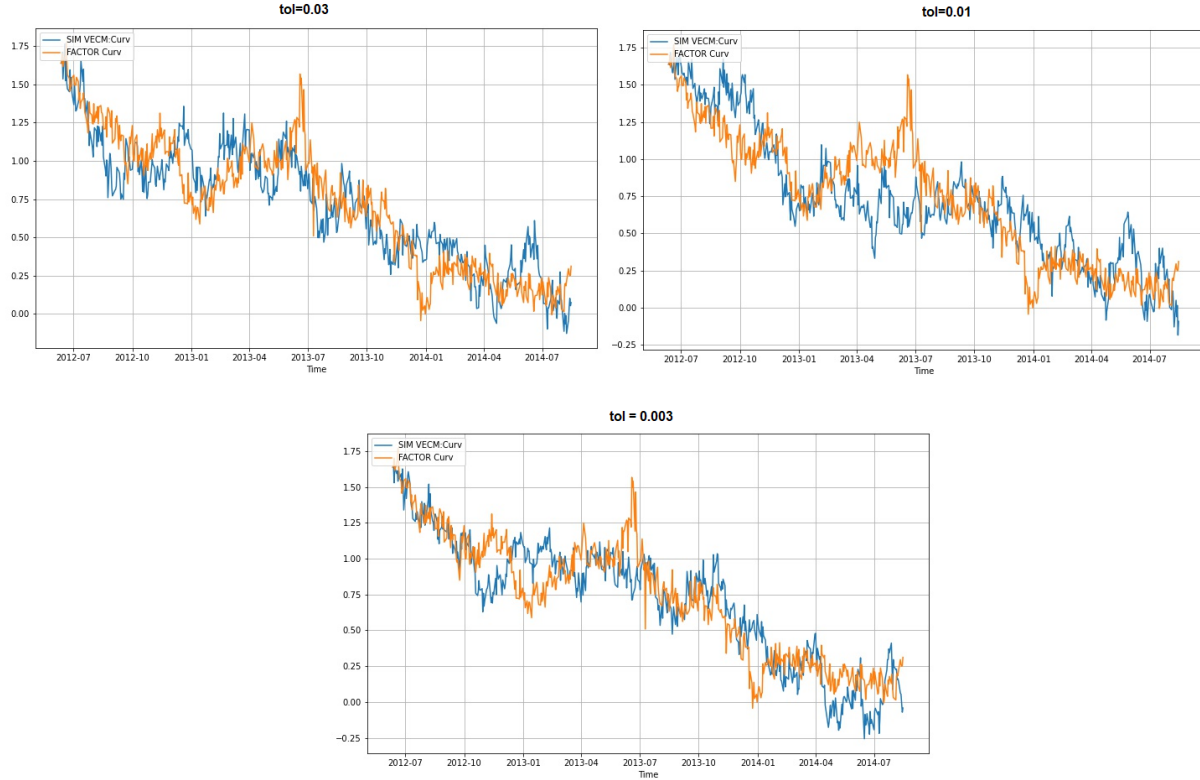
Nota. La media estimada corresponde a la media muestral obtenida con los datos de predicción para el valor en T en cada simulación. La std muestral corresponde a la desviación estándar de ese estimador.

Tabla 4.13: Resultados de VECM que incluye el factor 3 o curvatura y el driver MBS con datos diarios desde 12/06/2012 al 15/08/2014.

Resultados Predicción C_{550}	$tol = 0.03$	$tol = 0.01$	$tol = 0.003$
Lag VECM (BIC)	2	2	2
Número de Simulaciones	100	327	3275
Media Estimada Factor 3 (%)	0.066	0.061	0.062
Sesgo Factor 3	(-0.246)	(-0.251)	(-0.250)
Media Estimada Driver MBS (Dur)	0.483	0.471	0.470
Sesgo Driver MBS	0.031	0.019	0.018
Std Muestral Factor 3	0.013	0.005	0.001
Std Muestral Driver MBS	0.023	0.011	0.004
Tiempo Ejecución	16 s.	50 s.	446 s

La Figura 4.18 muestra la trayectoria del proceso del factor 3 durante la última simulación para $tol = 0.03$, $tol = 0.01$ y $tol = 0.003$.

Figura 4.18: *Trayectoria del Factor 3-Caso C (VECM): Muestra Real (Factor Curvatura) y última simulación (VECM Curvatura).*



Observación. Podemos observar que el modelo VAR necesitó un 300% más de simulaciones que el modelo VECM para alcanzar la tolerancia pretendida en los casos $tol = 0.01$ y $tol = 0.003$. De la misma manera, se necesitó un porcentaje similar de tiempo de ejecución.

El test ADF asumiendo una regresión con constante y tendencia (ct) para el proceso $\beta\tilde{X}_t$ arroja los resultados descriptos en La Tabla 4.14. Recordar que \tilde{X}_t es el proceso VAR formado por el factor 3 y el driver MBS.

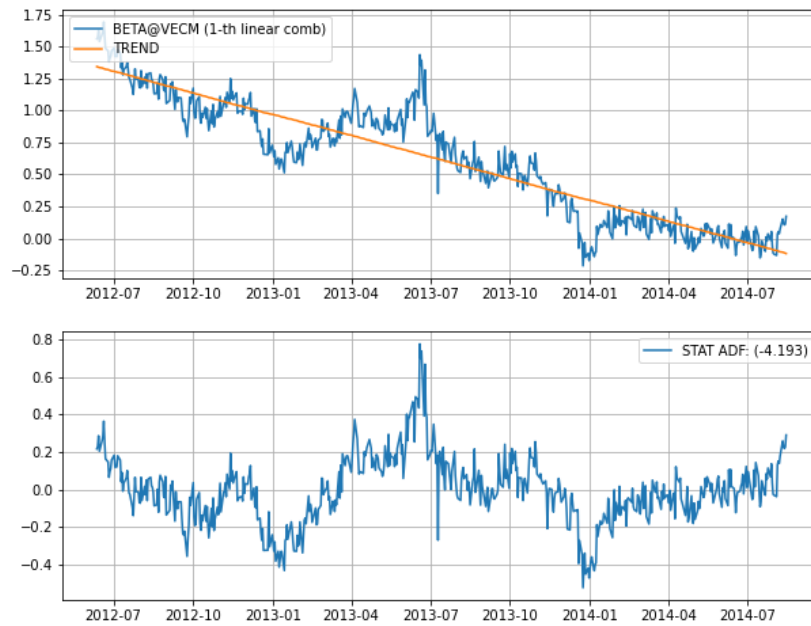
Es decir, el estadístico del test ADF para el proceso $\beta\tilde{X}_t$ es (-4.233) con p -valor igual a 0.004. Entonces, podemos rechazar la hipótesis de la raíz unitaria para este proceso estacionario con tendencia con un nivel de significancia del 1%.

Finalmente, en los gráficos de las Figuras 4.19 podemos observar el comportamiento de la serie $\beta\tilde{X}_t$ junto con la trayectoria de la tendencia obtenida por cuadrados mínimos para esa serie.

El test ADF sin asumir término de constante ni tendencia (nc) para la serie $\{\beta\tilde{X}_t - (1.344 - 0.003t)\}_t$ se describe en la Tabla 4.15.

Tabla 4.14: Test ADF (ϵ_t) para VECM - Caso C

<i>Estadístico del Test</i>	(-4.233)		
<i>p-valor</i>	0.004		
<i>Valores críticos del test estadístico en los niveles</i>	1 % (-3.975)	5 % (-3.419)	10 % (-3.132)

Figura 4.19: Trayectoria de $\beta\tilde{X}_t$: Tendencia (Factor Curvatura-MBS) y su serie de diferencias.

Es decir, el estadístico del test viene dado por (-4.193), lo que nos permite afirmar que la hipótesis de raíz unitaria es rechazada para ese proceso con un nivel de significancia del 1 %.

Tabla 4.15: Test ADF (nc) - Caso C

Estadístico del test	(-4.193)		
<i>p</i> -valor	3×10^{-5}		
Valores críticos del estadístico en los niveles	1 % (-2.570)	5 % (-1.941)	10 % (-1.616)

4.6.4. Resumen de resultados

En las siguientes Tablas 4.16, 4.17 y 4.18 se presenta un resumen de los resultados obtenidos previamente para cada caso de estudio.

Tabla 4.16: Caso A: Resumen resultados.

Niveles de Tolerancia	<i>tol</i> = 0.03		<i>tol</i> = 0.01		<i>tol</i> = 0.003	
Modelo Predicción	VAR	VECM	VAR	VECM	VAR	VECM
Número de Simulaciones	100	100	327	189	3440	2128
Std Muestral Factor 1	0.019	0.013	0.010	0.007	0.003	0.002
Std Muestral Driver MBS	0.011	0.023	0.006	0.019	0.002	0.006

Tabla 4.17: Caso B: Resumen resultados.

Niveles de Tolerancia	<i>tol</i> = 0.03		<i>tol</i> = 0.01		<i>tol</i> = 0.003	
Modelo Predicción	VAR	VECM	VAR	VECM	VAR	VECM
Número de Simulaciones	100	100	699	523	7468	6763
Std Muestral Factor 2	0.028	0.026	0.010	0.009	0.003	0.003
Std Muestral Driver USD/JPY	0.015	0.012	0.005	0.005	0.002	0.001

Tabla 4.18: Caso C: Resumen resultados.

Niveles de Tolerancia	$tol = 0.03$		$tol = 0.01$		$tol = 0.003$	
Modelo Predicción	VAR	VECM	VAR	VECM	VAR	VECM
Nro de Simulaciones	131	100	1222	327	13413	3275
Std Muestral Factor 3	0.030	0.013	0.010	0.005	0.003	0.001
Std Muestral Driver MBS	0.011	0.023	0.003	0.011	0.001	0.004

Capítulo 5

Conclusiones y futuras líneas de investigación

5.1. Conclusiones

En todos los estudios llevados a cabo anteriormente, observamos un mejor desempeño de nuestro modelo que del modelo VAR sin considerar cointegración. Para el caso A, observamos que nuestro modelo VECM es más eficiente que el modelo VAR original en términos de la definición 4.4.1 para la estimación del factor de nivel. De hecho, asumiendo una tolerancia de 0.03, la desviación del estimador es un 42 % mayor para las estimaciones del VAR que para las estimaciones del VECM alcanzada en el mínimo número de iteraciones. Para una tolerancia de 0.01, encontramos que el modelo VAR necesitó realizar un 73 % de simulaciones más que el modelo VECM y para la tolerancia de 0.003, necesitó un 62 % más de simulaciones para lograr la tolerancia. Se obtuvieron resultados similares para el caso de predicción de dos pasos como se puede ver en el Apéndice D. Además, si comparamos las gráficas de la trayectoria correspondiente a la última simulación (Figuras 4.9 y 4.10), observamos un mejor ajuste en el caso de nuestro modelo VECM. Finalmente, encontramos una relación de cointegración dada por $\beta = [1, -0.501]$. Esto nos dice que es suficiente restar la mitad del driver MBS al primer factor para obtener una serie estacionaria. Específicamente, una serie estacionaria con tendencia. En otras palabras, la variable de nivel L_t y el driver MBS_t tienen una tendencia estocástica común y el proceso $L_t - 0.501 \cdot MBS_t$ es estacionario, salvo una tendencia determinista.

Para el caso B, observamos que nuestro modelo VECM es más eficiente que el modelo VAR original en términos de la Definición 4.4.1 para la estimación del factor de pendiente. De hecho, asumiendo una tolerancia de 0.03, la desviación del estimador es 8,5 % mayor para las estimaciones de VAR que para la estimación del VECM alcanzada en el número mínimo de iteraciones. Para una tolerancia de

0.01, encontramos que el modelo VAR necesitó realizar un 34 % de simulaciones más que el modelo VECM y para la tolerancia de 0.003, necesitó un 10 % más de simulaciones para lograr las tolerancias. Se obtuvieron resultados similares para el caso de predicción de dos pasos como se puede ver en el Apéndice D. Además, si comparamos las gráficas de la trayectoria correspondiente a la última simulación (Figuras 4.13 y 4.14), observamos un mejor ajuste en el caso de nuestro modelo VECM. Finalmente, encontramos una relación de cointegración dada por $\beta = [1, -2.1365]$. Esto nos dice que es suficiente restar el doble del driver USD/JPY del primer factor para obtener una serie estacionaria. Específicamente, una serie estacionaria con tendencia. En otras palabras, la variable pendiente S_t y el driver USD/JPY $_t$ tienen una tendencia estocástica común y el proceso $S_t - 2.1365 \cdot \text{JPY}_t$ es estacionario, salvo una tendencia determinista.

Para el caso C, observamos que nuestro modelo VECM es más eficiente que el modelo VAR original en términos de la Definición 4.4.1 para la estimación del factor de curvatura. De hecho, asumiendo una tolerancia de 0.03 la desviación del estimador fue un 135 % mayor para las estimaciones de VAR que para la estimación de VECM. De hecho, en el caso del VAR fueron necesarias 131 simulaciones y en el caso del VECM la tolerancia se alcanzó en el número mínimo de iteraciones (igual a 100). Para una tolerancia de 0.01, encontramos que el modelo VAR necesitó realizar un 274 % de simulaciones más que el modelo VECM y para la tolerancia de 0.003, necesitó un 310 % más de simulaciones para lograr las tolerancias. Se obtuvieron resultados similares para el caso de predicción de dos pasos, como se puede observar en el Apéndice D. Además, si comparamos las gráficas de la trayectoria correspondiente a la última simulación (Figuras 4.17 y 4.18), observamos un mejor ajuste en el caso de nuestro modelo VECM. Finalmente, encontramos una relación de cointegración dada por $\beta = [1, -0.3084]$. Esto nos dice que basta restar el 31 % del driver MBS del primer factor para obtener una serie estacionaria. Específicamente, una serie estacionaria con tendencia. En otras palabras, la variable de curvatura C_t y el driver MBS $_t$ tienen una tendencia estocástica común y el proceso $C_t - 0.3084 \cdot \text{MBS}_t$ es estacionario, salvo una tendencia determinista.

5.2. Futuras líneas de investigación

Como una continuidad del presente trabajo, se propone llevar a cabo el mismo análisis para un conjunto de datos más reciente.

Por otro lado, la metodología desarrollada aquí podría ser aplicada a otra curva de tasas de interés. Para ello, sería necesario realizar un análisis previo para decidir los drivers económicos que afectan a tal curva.

En esta línea podría ser de gran aporte encontrar casos o aplicaciones donde se aproveche la estructura vectorial de la serie a modelar y que la dimensión de tal serie sea superior a dos de manera de encontrar

una cointegración de un factor con más de un driver.

También consideramos que podría ser de interés realizar una estimación de las series obtenidas a través del modelo de Nelson y Siegel por algún otro método que permita estimar el parámetro λ de manera conjunta al resto de parámetros (ver Diebold et al. [20]).

Apéndice A

Modelos de tasa de interés

Los *modelos de la estructura a término* buscan describir la evolución de todas las tasas de interés spot para cada madurez correspondientes a la estructura a término. Esto se puede realizar especificando diversas variables de dicha estructura: la tasa short, la tasa forward, entre otras. En este apéndice comenzaremos dando el marco teórico financiero que servirá de sustento para describir los diversos modelos.

En las siguientes secciones daremos una clasificación detallada de los diferentes modelos ya desarrollados en la literatura que buscan explicar y analizar las diversas curvas de tasas de interés. La estructura de dicha clasificación sigue la línea de trabajo realizada en Saúl Gri [53].

A.1. Conceptos generales del modelado de tasas de interés

A.1.1. Proceso de precios de bonos y modelado general de tasa short

En esta sección plantaremos resultados propios de la teoría financiera para el modelado de tasas de interés. Estos serán de importancia tanto para definir y construir nuestro modelo como para comprender los ya existentes.

Siguiendo a Musiela and Rutkowski [46], consideremos:

- (Ω, \mathcal{F}, P) un espacio de probabilidad.
- $T^* > 0$ una fecha de horizonte fija para todas las actividades del mercado.
- $\mathcal{F} = \{\mathcal{F}_t\}_{t \geq 0}$ la filtración generada por el movimiento Browniano W_t subyacente.
- para cualquier madurez fija $T \leq T^*$, el proceso de precios $P_t = P(t, T)$, $t \in [0, T]$ es un proceso \mathcal{F}_t -adaptado y estrictamente positivo.

Supongamos que un proceso r que es \mathcal{F}_t -adaptado (con respecto a P) modela la tasa de interés short. Esto significa que la cuenta de mercado de moneda sigue el proceso

$$B(t) = \exp\left(\int_0^t r_s ds\right), \quad \text{para todo } t \in [0, T^*].$$

Luego, podemos dar la siguiente definición de un proceso libre de arbitraje.

Definición A.1.1. *El proceso adaptado $\{P(t, T)\}_{0 \leq t \leq T \leq T^*}$ es llamado un proceso de precios de bonos libre de arbitraje relativo a una tasa short r si satisface las siguientes condiciones:*

1. $P(T, T) = 1$;
2. Existe una medida de probabilidad P^* en (Ω, \mathcal{F}) equivalente a P tal que el proceso de bonos descontado

$$Z(t, T) \doteq P(t, T)/B(t)$$

es una martingala bajo P^* .

Cualquier medida de probabilidad P^* dada por la definición anterior se denomina una *medida de martingala para el proceso $\{P(t, T)\}_{t \in [0, T]}$* .

Supongamos ahora que la tasa de interés short r sigue el proceso de Itô bajo la probabilidad real P dado por:

$$dr_t = \mu_t dt + \sigma_t dW_t, \quad r_0 > 0, \quad (\text{A.1})$$

con $\mu_t = \mu(t, r)$ y $\sigma_t = \sigma(t, r)$ procesos estocásticos adaptados con valores en \mathbb{R} .

Proposición A.1.2. *Supongamos que la tasa de interés short r sigue el proceso de Itô bajo la probabilidad real P dado por (A.1). Sea $P(t, T)$ una familia de precios de bonos libre de arbitraje relativa a r . Para cualquier medida de martingala $P^* = P^\lambda$ de la definición (A.1.1), las siguientes afirmaciones son verdaderas:*

1. El proceso r satisface bajo P^λ

$$dr_t = (\mu_t + \sigma_t \lambda_t) dt + \sigma_t dW_t^\lambda.$$

2. Existe un proceso adaptado $b^\lambda(t, T)$ a valores en \mathbb{R} tal que

$$dP(t, T) = P(t, T)(r_t dt + b^\lambda(t, T) dW_t^\lambda).$$

En consecuencia, para cualquier madurez fija $t \in (0, T^*]$ tenemos:

$$\begin{aligned} P(t, T) &= P(0, T) W_t \eta_t \\ &= P(0, T) W_t \exp\left(\int_0^t b^\lambda(u, T) dW_u^\lambda - \frac{1}{2} \int_0^t |b^\lambda(u, T)|^2 du\right), \end{aligned}$$

con η_t la derivada Radon-Nikodým del proceso b^λ .

Es importante notar que el proceso $b^\lambda(t, T)$ no depende de la elección de λ . Esto es una consecuencia inmediata del teorema de Girsanov. Luego, podemos formular la siguiente definición.

Definición A.1.3. *El proceso $b(t, T) = b^\lambda(t, T)$ es denominado la volatilidad de un bono cupón cero con madurez T .*

A.1.2. Expectativas, prima de riesgo y convexidad

En esta sección buscaremos conectar las ideas de expectativas y prima de riesgo con el proceso de tasas de interés. Además, debido a que la volatilidad en torno a las expectativas de los participantes del mercado afectan la forma de la curva de rendimiento, es importante comprender el “efecto de convexidad” en dicha curva.

Expectativas

La palabra expectativa implica incertidumbre. Los inversores podrían esperar que la tasa a un año sea del 10 %, pero también saben que hay una probabilidad de que resulte ser del 8 % o del 12 %.

Las expectativas pueden hacer que la estructura temporal asuma cualquiera de una gran variedad de formas. Sobre horizontes cortos, uno puede imaginar que la comunidad financiera tiene visiones específicas sobre el futuro de la tasa de corto plazo. Durante horizontes largos, el camino de las expectativas puede no ser tan específico. Por lo tanto, las previsiones pueden ser muy útiles para describir el nivel y la forma de la estructura temporal en horizontes cortos de tiempo pero sólo el nivel de las tasas en horizontes muy largos.

Esta conclusión tiene importantes implicaciones a la hora de extraer expectativas de las tasas de interés observadas para las técnicas de ajuste de curvas no basadas en modelos de estructura temporal, y para el uso de modelos libres de arbitraje de la estructura de plazos como veremos en la próxima sección.

Un aspecto clave en el modelado de tasas de interés es la conexión entre el concepto económico de ausencia de arbitraje y la propiedad matemática de existencia de una medida de probabilidad para la cual la familia de precios de bonos de cualquier vencimiento forman una martingala. Esto se encuadra en lo que se denominan “teorías de las expectativas” y que detallamos brevemente a continuación:

Consideremos $\{P(t, T)\}_{t \in [0, T]}$ un proceso de precios de bonos libre de arbitraje relativo a la tasa short r como el dado en la Definición A.1.1. Sea Q la medida de probabilidad real y supongamos que

$$P(t, T) = E_Q \left(e^{-\int_t^T r_s ds} | \mathcal{F}_t \right). \quad (\text{A.2})$$

Notar que estamos asumiendo la tasa short r como un proceso estocástico más que determinista $r = \{r_t\}_{t \in [0, T]}$.

La última igualdad se conoce tradicionalmente como una *hipótesis de expectativas locales*, o una *hipótesis de expectativas de riesgo neutral*. El término “expectativas locales” se refiere al hecho de que en virtud de esta igualdad, *el precio del bono actual es igual al valor esperado, bajo la probabilidad real, del precio del bono en el siguiente período (infinitesimal), descontado a la tasa short actual*.

Varios autores han analizado la validez de las diversas formas de hipótesis de expectativas bajo la probabilidad real, por lo general en el marco de un enfoque de equilibrio general, concepto que veremos en la próxima sección (ver Cox et al. [13], Fama [26], Campbell [8], Longstaff [42], Stigler [56] y McCulloch [45]).

Volatilidad y convexidad

Consideremos nuevamente el proceso de precios de bonos libre de arbitraje $\{P(t, T)\}_{t \in [0, T]}$ como en la Definición A.1.1. Se puede ver que dicho proceso cumple

$$P(t, T) = E_{P^*} \left(e^{-\int_t^T r_s ds} \mid \mathcal{F}_t \right), \quad \forall t \in [0, T],$$

donde P^* es la medida de martingala correspondiente al proceso.

Utilizando la *Desigualdad de Jensen condicional* para la función convexa $g(x) = e^{-x}$ obtenemos:

$$E_{P^*} \left(e^{-\int_t^T r_s ds} \mid \mathcal{F}_t \right) \geq e^{-E_{P^*}(\int_t^T r_s ds \mid \mathcal{F}_t)}.$$

Esta desigualdad dice que el precio del bono libre de arbitraje en t es mayor que el precio del bono en t obtenido de descontar a la tasa esperada durante el período $[t, T]$. Es decir, que la tasa obtenida para el período $[t, T]$ es menor que la tasa esperada en dicho período.

El hecho de incorporar incertidumbre o *volatilidad* a la tasa short r genera que la tasa de interés obtenida a partir del precio del bono sea menor que la tasa de interés esperada. Esta diferencia se denomina el *efecto de convexidad*.

En general se observa que dejando fijos los demás parámetros la convexidad aumenta con la madurez. Además, existe evidencia empírica de que las tasas de interés a corto plazo fluctúan más que las tasas a largo plazo; esto se observa en parte debido a que la forma típica de la estructura temporal de las volatilidades de los rendimientos tiene pendiente negativa. Luego, si bien la convexidad aumenta con la madurez, ésta es compensada por la reducción de la volatilidad.

Concluimos que tanto las expectativas como la convexidad son factores a tener en cuenta para entender las formas que puede asumir la estructura temporal de tasas de interés. Estas características hacen que la construcción de un modelo fiable para el comportamiento estocástico de la estructura temporal de tasas de interés sea una tarea de gran complejidad.

Prima de riesgo

Para comprender el significado de prima de riesgo, podemos comenzar dando una idea intuitiva de la misma: la prima de riesgo es la compensación que los inversores exigen por asumir el riesgo de que los rendimientos no evolucionen como ellos esperan.

En un marco de neutralidad al riesgo, es decir, si se supusiera que no existe riesgo en la inversión de bonos, los rendimientos de los mismos deberían ser iguales al valor esperado de las tasas short futuras. Como detallamos anteriormente, esto es lo que supone la hipótesis de expectativas de la estructura temporal que establece que los movimientos en las tasas de largo plazo se deben a movimientos en las tasas short futuras esperadas. Cualquier término o prima de riesgo se asumen constantes en el tiempo.

En este contexto, la hipótesis de expectativas, que sigue siendo uno de los pilares de muchos análisis macroeconómicos casuales y formales, es válido. Sin embargo, parece probable que los bonos, que proporcionan una rentabilidad incierta, sean propiedad de los mismos inversores que también demandan una gran prima de riesgo como compensación por mantener los activos riesgosos.

Por otra parte, según lo sugerido por muchos tests estadísticos en la literatura, estas primas de riesgo de los bonos parecen variar con el tiempo, lo que contradice la suposición de neutralidad al riesgo. Por ejemplo, Campbell and Shiller [9] y Cochrane and Piazzesi [12] proporcionan evidencia reciente sobre el fracaso de la hipótesis de expectativas.

Para modelar estas primas, Ang and Piazzesi [2] y Rudebusch and Wu [52, 51] especifican “precios de riesgo” variables en el tiempo que traducen una unidad de factor de volatilidad en una prima de plazo. Esta variación en el tiempo se modela utilizando indicadores del ciclo económico, como la pendiente de la curva de rendimiento o medidas de la actividad real. Sin embargo, Diebold et al. [20] sugieren que la importancia de las desviaciones estadísticas de la hipótesis de expectativas puede depender de la aplicación.

En la literatura financiera hay dos enfoques básicos para el modelado de las primas de riesgo variables en el tiempo: cantidades de riesgo variables en el tiempo o precios de riesgo variables en el tiempo. La amplia literatura sobre volatilidad estocástica toma el primer enfoque, lo que permite la variabilidad de los factores a cambiar con el tiempo. En contraste, la representación financiera y Gaussiana afín libre de arbitraje (por ejemplo, Ang and Piazzesi [2]) toma el último enfoque, especificando precios de riesgo variables en el tiempo, los cuales traducen una unidad de factor de volatilidad en una prima de riesgo.

Para intentar captar esta prima de riesgo, supongamos que la tasa de interés short r sigue el proceso de Itô bajo la probabilidad real P dado por:

$$dr_t = \mu_t dt + \sigma_t dW_t, \quad r_0 > 0, \quad (\text{A.3})$$

donde $\{W_t\}_{0 \leq t \leq T}$ es un movimiento Browniano d -dimensional bajo P con $\mu_t = \mu(t, r)$ y $\sigma_t = \sigma(t, r)$ procesos estocásticos adaptados con valores en \mathbb{R} y \mathbb{R}^d , respectivamente.

Sea $P^* = P^\lambda$ una medida de martingala arbitraria equivalente a P . Entonces, por lo visto en (A.2), podemos definir el precio de un bono $P(t, T)$ de la siguiente manera:

$$P(t, T) = E_{P^*} \left(e^{-\int_t^T r_s ds} \middle| \mathcal{F}_t^{W^\lambda} \right), \quad \forall t \in [0, T]. \quad (\text{A.4})$$

De la Proposición A.1.2 se sigue que el precio del bono $P(t, T)$ satisface, bajo la probabilidad P ,

$$\begin{aligned} dP(t, T) &= P(t, T)((r_t - \lambda_t \cdot b^\lambda(t, T))dt + b^\lambda(t, T) \cdot dW_t). \\ &= P(t, T)(\mu_t^P dt + b^\lambda(t, T) \cdot dW_t), \end{aligned} \quad (\text{A.5})$$

donde $\mu_t^P \doteq r_t - \lambda_t \cdot b^\lambda(t, T)$ representa el retorno del bono en la medida real P y $\lambda_t = \frac{r_t - \mu_t^P}{b^\lambda(t, T)}$. La igualdad (A.5) se deduce del hecho que el movimiento Browniano W^λ cumple con la siguiente relación como resultado del teorema de Girsanov's:

$$W_t^\lambda = W_t - \int_0^t \lambda_u du, \quad \forall t \in [0, T^*],$$

que en su forma diferencial, es equivalente a:

$$dW_t^\lambda = dW_t - \lambda_t dt.$$

Esto significa que los retornos instantáneos de mantener el bono difieren, en general, de la tasa de interés short r . En la literatura financiera, este término adicional se conoce comúnmente como la *prima de riesgo* o *el precio de mercado para el riesgo de la tasa de interés*. Es decir, recordemos que r es la tasa instantánea de retorno de una inversión libre de riesgo, por lo que la diferencia $\mu^P - r$ representa una diferencia en los retornos. Esto nos dice cuán mejor es nuestro rendimiento con respecto al caso libre de riesgo, es decir, con respecto a poner nuestro dinero en una cuenta bancaria sin riesgo. Cuando dividimos esta cantidad por b^λ , estamos dividiendo por la cantidad de riesgo al que estamos expuestos, medida por la volatilidad de los precios de bonos b^λ . Esta es la razón por la que λ se conoce como “precio de mercado de riesgo”. Luego, si los supuestos se refieren sólo a la ausencia de arbitraje entre los securities primarios y derivados (es decir, si uno se coloca en un marco de equilibrio parcial), no se puede identificar la prima de riesgo.

En resumen, tenemos un cierto grado de libertad; si la tasa a corto plazo r está dada por (A.3), entonces cualquier medida de probabilidad P^* equivalente a P se puede utilizar formalmente para construir una familia libre de arbitraje de precios de bonos a través de la fórmula (A.4). Nótese, sin embargo, que si la medida de probabilidad real P se utiliza para definir el precio del bono a través de la fórmula (A.4), las primas de riesgo se desvanecen.

A.2. Modelos de un factor: tasa short

En esta sección nos centraremos en modelos de estructura a término contruidos especificando sólo el comportamiento de la tasa short que denotaremos r . También nos referiremos a esta tasa como la tasa de interés de corto plazo.

Comenzaremos detallando los modelos de tasa short abordando dos tipos de modelos que sirven de base para comprender modelos más sofisticados. Estos se denominan modelos de equilibrio y modelos libres de arbitraje de la tasa short.

Los modelos libres de arbitraje comienzan con supuestos sobre el comportamiento estocástico de una o varias tasas de interés y sobre un precio de mercado de riesgo específico y obtienen el precio de todos los contingent claims asumiendo que no existen oportunidades de arbitraje en el mercado. En otros términos, no hay una estrategia financiera libre de riesgo con costo cero que deba dar con certeza un retorno positivo. Por el contrario, los modelos de equilibrio parten de una descripción de la economía y derivan la estructura temporal de tasas de interés, la prima de riesgo y los precios de otros activos de manera endógena, en el supuesto de que el mercado está en equilibrio. Pero la distinción es sutil, ya que los modelos de equilibrio deben ser libre de arbitraje (de lo contrario, la economía no estaría en equilibrio), y por otro lado, existen modelos llamados “libres de arbitraje” en los que se permiten oportunidades de arbitraje.

A.2.1. Modelos de equilibrio

Los modelos de equilibrio derivan de algún modelo de equilibrio general de la economía. Para ellos, una comprensión de las relaciones entre los supuestos del modelo y de la forma de la estructura temporal es importante con el fin de hacer suposiciones razonables en primer lugar. Para los modelos libres de arbitraje, la comprensión de estas relaciones revela los supuestos implícitos en el mercado a través de la estructura temporal observada.

Los modelos de equilibrio determinan el proceso neutral al riesgo de la tasa de interés short r a partir de:

- la hipótesis sobre el proceso de tasas de interés real;
- las hipótesis sobre la prima de riesgo demandada por el mercado para asumir el riesgo de las tasas de interés.

A través de los modelos de equilibrio se puede hacer una conexión entre la dinámica de la tasa short y la prima de riesgo con la forma de la estructura temporal. Si bien esto puede no dar respuesta a preguntas en niveles macroeconómicos, es más ambicioso que la derivación de restricciones de arbitraje en precios de bonos y derivados, dados los precios de un conjunto de bonos subyacentes.

Suponiendo que la economía tiende hacia algún equilibrio basado en factores fundamentales como la productividad del capital, la política monetaria a largo plazo, entre otros, entonces las tasas short estarán caracterizadas por *reversión a la media*. Luego, asumiendo que dichas tasas se mueven como un resultado de algunas noticias o *shock* del sistema económico, reversión a la media significa que el efecto de ese shock eventualmente se disipa.

Entre los modelos clásicos más estudiados que podemos encontrar con estas características se encuentran, cronológicamente: el modelo de Vasicek (Vasicek [61], 1977), el modelo de Dothan (Dothan [21], 1978), el modelo de Cox, Ingersoll y Ross (Cox et al. [14], 1985) o comúnmente llamado modelo CIR y el modelo de Black-Karasinski (Black and Karasinski [5], 1991).

Podemos destacar que estos modelos tienen la característica de ser endógenos pues la estructura temporal de tasas es un output en lugar de ser un input del modelo.

A continuación detallaremos dos de los modelos previamente mencionados: el modelo de Vasicek y el modelo CIR.

Modelo de Vasicek

El enfoque pionero propuesto por Vasicek está basado en la definición de la dinámica de la tasa short r bajo la probabilidad real. La derivación de un precio libre de arbitraje para cualquier derivado de tasas de interés se desprende del uso de los argumentos básicos de Black y Scholes (Black and Scholes [6]), teniendo en cuenta la característica de que las tasas de interés no son un instrumento negociable. La construcción de un portfolio localmente libre de riesgo adecuado, como en Black y Scholes, da lugar a la existencia de un proceso estocástico que sólo depende del tiempo actual y la tasa spot instantánea y no de los vencimientos de los activos que constituyen el portfolio. Tal proceso, que denominamos previamente “precio de mercado del riesgo”, define un cambio de medida de Girsanov del mundo real al mundo neutral al riesgo también en casos de dinámicas más generales que la de Vasicek. La dinámica neutral al riesgo del modelo de Vasicek viene dada por el siguiente proceso:

$$dr_t = k(\theta - r_t)dt + \sigma dW_t, \quad (\text{A.6})$$

donde θ denota el *long-run value o tendencia central* de la tasa short y k es una constante positiva que denota la velocidad de reversión a la media. Como antes, $\{W_t\}$ es un movimiento Browniano bajo la medida de riesgo neutral.

Esta dinámica tiene algunas peculiaridades que hacen atractivo al modelo. La ecuación es lineal y se puede resolver explícitamente, la distribución de la tasa short es Gaussiana, y tanto las expresiones como las distribuciones de muchas cantidades útiles relacionadas al mundo de las tasas de interés se obtienen fácilmente. Además, la naturaleza endógena del modelo es clara. Debido a que el precio del bono

$$P(t, T) = E(e^{-\int_t^T r_s ds} | \mathcal{F}_t) \quad (\text{A.7})$$

se puede calcular como una simple expresión que depende de k , θ , σ y r_t , una vez que la función $T \mapsto P(t, T; k, \theta, \sigma, r_t)$ es conocida, queda determinada la curva de tasas de interés completa en el tiempo t . Esto significa que, si $t = 0$ es el tiempo inicial, la curva de tasas de interés inicial es un output del modelo, dependiendo de los parámetros k , θ , σ en la dinámica (y en la condición inicial r_0). Sin embargo, este modelo tiene algunas desventajas. Por ejemplo, existe una probabilidad positiva de que las tasas asuman valores negativos.

Debido a que el proceso de (A.6) es un proceso neutral al riesgo, el drift combina tanto las expectativas de tasas de interés como la prima de riesgo. Por otra parte, los precios de mercado no dependen de cómo el drift neutral al riesgo se divide a través de sus dos fuentes. Sin embargo, para entender si los parámetros de un modelo tienen sentido, es útil hacer suficientes suposiciones para separar el drift y la prima de riesgo (ver sección A.1.2).

Modelo de Cox-Ingersoll-Ross

El modelo de Vasicek supone que la volatilidad de la tasa short es independiente del nivel de dicha tasa. Es bastante probable que esto no se cumpla en niveles extremos de la tasa short. Los períodos de inflación y tasas altas son naturalmente inestables y, como consecuencia, la volatilidad de la tasa short tiende a ser elevada. Además, cuando la tasa de corto plazo es muy baja, su volatilidad está limitada por el hecho de que las tasas de interés no pueden declinar muy por debajo de cero.

Argumentos económicos de este tipo han conducido a especificar la volatilidad de la tasa short como una función creciente de dicha tasa. Un modelo con esta característica es el modelo de Cox-Ingersoll-Ross (CIR) cuya dinámica neutral al riesgo viene dada por:

$$dr_t = k(\theta - r_t)dt + \sigma\sqrt{r_t}dW_t. \quad (\text{A.8})$$

En este modelo, el parámetro σ es constante, pero la volatilidad no lo es: la volatilidad anualizada es igual a $\sigma\sqrt{r_t}$ y, por lo tanto, aumenta con el nivel de la tasa short. En este caso, el parámetro σ es a menudo llamado *yield volatility*.

A.2.2. Modelos libre de arbitraje

En el ámbito profesional, muchas veces se desea suponer que algún conjunto de securities es valorado de manera justa para luego poder valorar y cubrir derivados de securities con el propósito de generar mercados o para trading propio. Debido a que por construcción los modelos libres de arbitraje ajustan los precios de muchos securities negociados, estos modelos son ideales para el propósito de valorar derivados dados los precios de los securities subyacentes.

Los modelos libres de arbitraje determinan el proceso neutral al riesgo a partir de:

- la estructura a término inicial;
- hipótesis sobre el proceso de tasas de interés real para la tasa short.

La interpolación por medio de modelos libres de arbitraje puede ser muy superior a otros métodos de ajuste de curvas, desde interpolación lineal a enfoques más sofisticados como splines cúbicos. La superioridad de los modelos libres de arbitraje surge de que ellos se construyen sobre la base de un razonamiento económico y financiero. Esta presunta superioridad depende fundamentalmente de la validez de los supuestos incorporados en dichos modelos. Una hipótesis pobre de volatilidad, por ejemplo, implicaría una mala estimación del efecto de convexidad y podría hacer que un modelo libre de arbitraje resulte peor que una técnica menos sofisticada financieramente.

Que un modelo coincida con los precios de mercado no implica necesariamente que proporcione valoraciones justas y coberturas precisas de derivados de securities. El argumento para ajustar modelos a precios de mercado es que una buena cantidad de información sobre el comportamiento futuro de las tasas de interés se incorpora en los precios de mercado, y, por tanto, un modelo ajustado a los precios captura el comportamiento de la tasa de interés. Si bien este es un argumento perfectamente razonable, señalamos dos advertencias:

- un modelo que no brinde buenos resultados no puede ser rescatado por la calibración para que coincida con los precios de mercado. Si, por ejemplo, la suposición de shift paralelo no es una descripción suficientemente buena de la realidad para la aplicación en cuestión, el hecho de añadir un time-dependent drift a un modelo parallel shift para que coincida con un conjunto de precios de bonos no hará que el modelo sea adecuado para esa aplicación.
- el argumento para ajustar a precios de mercado supone que esos precios de mercado son justos en el contexto del modelo. Hay, sin embargo, muchas situaciones en las que los securities particulares, clases particulares de securities, o determinados rangos de vencimiento de los securities han sido distorsionados debido a los desequilibrios de oferta y demanda, los impuestos, las diferencias de liquidez y otros factores no relacionados con los modelos de tipos de interés. En estos casos, el ajuste a precios de mercado hará un modelo menos fiable atribuyendo estos factores externos al proceso de tasa de interés.

Modelo Ho-Lee

Un ejemplo de modelo libre de arbitraje es el modelo Ho-Lee (ver Ho and Lee [32]) que posee un drift dependiente del tiempo. La dinámica del proceso neutral al riesgo para la tasa short r en dicho modelo es:

$$dr_t = \lambda_t dt + \sigma dW_t; \quad (\text{A.9})$$

donde el drift λ depende del tiempo y $\{W_t\}$ es un movimiento Browniano bajo la medida de riesgo neutral. Esta dinámica continua de la tasa short se obtiene, de hecho, de un límite apropiado de un modelo binomial (ver Ji and Yin [34]).

Este modelo toma como dada la estructura temporal de tasas de interés *completa* y deriva el movimiento estocástico posterior de la estructura a término de manera tal que dicho movimiento sea libre de arbitraje. La curva de tasas resultante a partir de este modelo es tal que se ajusta a todas las tasas que son input en el modelo.

Si el propósito de un modelo es valorar bonos o swap relativos a otros, entonces tomar como dados un gran número de precios de bonos o de swaps es claramente inapropiado: por construcción, los modelos libres de arbitraje concluyen que todos estos precios de bonos o de swaps son justos relativos a otro. Los inversores que deseen elegir entre securities, market makers que buscan captar valor seleccionando estratégicamente securities de cobertura, o los traders que buscan ganar desde esos mispricings temporales deben confiar, por lo tanto, en *modelos de equilibrio*.

A.3. Modelos multifactoriales

Los modelos de una variable asumen que los cambios en toda la estructura temporal de tasas de interés se pueden explicar por los cambios de una única tasa. Los modelos difieren en cómo esa tasa impacta sobre la estructura temporal, pero en todos los modelos, las tasas de todos los términos están perfectamente correlacionados. Según estos modelos, conocer el cambio en alguna tasa es suficiente para predecir perfectamente el cambio en cualquier otra. Pero, en general, esta propiedad contradice el comportamiento real de la curva de rendimientos. Sin embargo, para algunos fines puede ser apropiado un análisis de un factor.

El modelado de la evolución de tasas de interés a través de la tasa short instantánea tiene algunas ventajas, sobre todo la gran libertad que uno tiene en la elección de las dinámicas relacionadas. Por ejemplo, para los modelos de tasas short de un factor se puede elegir libremente el drift y el coeficiente de volatilidad instantánea en la dinámica de difusión relacionada como se considere conveniente, sin restricciones generales.

Sin embargo, por otro lado, los modelos de tasas short suelen tener varios inconvenientes. Por ejemplo, es difícil obtener una calibración exacta a la curva inicial de factores de descuento y una buena comprensión de la estructura de covarianza de las tasas forward, especialmente para los modelos que no son manejables analíticamente.

Finalmente, en el ámbito profesional de renta fija se suele estar expuesto al riesgo de la remodelación de la estructura temporal y los modelos de un factor, por lo general, resultan inadecuados.

En esta sección, seguiremos la bibliografía desarrollada en Duffie [23], Musiela and Rutkowski [46], Duffie and Kan [24] y Diebold and Rudebusch [19] para describir los modelos multifactoriales más

estudiados y utilizados.

Si bien la terminología es diversa en el marco multifactorial diremos que un modelo “multifactor” es tal que la tasa short tiene la forma $r_t = R(X_t, t)$, $t \geq 0$, con $X = \{X_t\}_{t \geq 0}$ un proceso de Itô en algún subconjunto $D \subseteq \mathbb{R}^k$ al que llamaremos “espacio de estados” y tal que X_t es solución de una ecuación diferencial estocástica de la forma

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t. \quad (\text{A.10})$$

Las funciones dadas R , μ , y σ de $D \times [0, \infty)$ en \mathbb{R} , \mathbb{R}^k , y $\mathbb{R}^{k \times d}$, respectivamente, satisfacen suficientes regularidades técnicas (ver Duffie [23]) para garantizar que (A.10) tenga solución única y que la estructura a término (A.2) esté bien definida. El proceso $\{W_t\}$ es un movimiento Browniano d -dimensional bajo la probabilidad de riesgo neutral.

Diremos que X_t es un “vector de estados”. Luego, una interpretación del rol de las “variables de estado” dependerá de cada modelo. Por ejemplo, en un modelo de equilibrio algunos elementos del vector de estado están a veces *latentes*, es decir, no observables para el modelador, excepto cuando ellos puedan ser inferidos desde precios que dependen de los niveles de X_t . Otro acercamiento es tomar algunas o todas las variables de estado como variables observables directamente, tales como determinantes macroeconómicos del ciclo económico e inflación, pues se piensa que desempeñan algún papel en la determinación de la estructura temporal.

A.3.1. Modelos afines

Una subclase importante de modelos multifactor es la de los *modelos afines de la estructura a término* para los cuales el proceso de estados $X = \{X_t\}_{t \geq 0}$ es un proceso estocástico *afín*, definido a través de (A.10) con

$$\mu(x, t) = K_0 + K_1 x,$$

para algún $K_0 \in \mathbb{R}^k$ y $K_1 \in \mathbb{R}^{k \times k}$, y tal que para cada i y j en $\{1, \dots, k\}$,

$$(\sigma(x, t)\sigma(x, t)^T)_{ij} = H_{0ij} + H_{1ij} \cdot x,$$

para $H_{0ij} \in \mathbb{R}$ y $H_{1ij} \in \mathbb{R}^k$. Se puede permitir también que las funciones coeficientes $H = (H_0, H_1)$ y $K = (K_0, K_1)$ dependan de t , aunque ignoraremos esto para simplificar notación.

Dado los coeficientes (H, K) , un espacio de estados natural $D \subset \mathbb{R}^k$ para este modelo es tal que se cumpla $(\sigma(x, t)\sigma(x, t)^T)_{ii} \geq 0$ para todo $x \in D$. Luego, dado H , elegimos el espacio de estados

$$D = \{x \in \mathbb{R}^k : H_{0ii} + H_{1ii} \cdot x \geq 0, i \in \{1, \dots, k\}\}.$$

Se puede ver (Duffie [23] ó Duffie and Kan [24]) que bajo ciertas condiciones en los coeficientes (H, K) se asegura la existencia de una única solución X para (A.10) valuado en D .

Ejemplo A.3.1. Un ejemplo es el modelo “CIR multifactor”, definido por:

$$dr_t^{(i)} = \kappa_i(\theta_i - r_t^{(i)})dt + C_i\sqrt{r_t^{(i)}}dW_t^{(i)}, \quad r_0^{(i)} > 0,$$

donde κ_i , θ_i y C_i son constantes positivas cumpliendo el mismo rol que κ , θ , y σ en el modelo CIR de un factor (A.8) y $\{W_t^{(i)}\}$ es un movimiento Browniano unidimensional. En este caso, tenemos que

$$K_{0_i} = \kappa_i\theta_i, \quad K_{1_{ij}} = \begin{cases} \kappa_i, & \text{si } i = j, \\ 0, & \text{si } i \neq j \end{cases}, \quad H_0 \equiv 0 \quad \text{y} \quad (H_{1_{ij}})_k = C_k^2.$$

Supongamos ahora que

$$R(x, t) = \rho_0 + \rho_1 \cdot x,$$

para coeficientes $\rho_0 \in \mathbb{R}$ y $\rho_1 \in \mathbb{R}^k$. Luego, dada una madurez fija T y debido a la independencia bajo la probabilidad de riesgo neutral de $W_t^{(1)}, \dots, W_t^{(k)}$, se obtiene una solución $P(X_t, t)$ para el precio en tiempo t de un bono cupón cero con madurez T de la forma exponencial afín

$$P(x, t) = e^{\alpha(t) + \beta(t) \cdot x},$$

con α y β deterministas (ver Duffie and Kan [24]).

En Duffie and Kan [24] se presenta el *affine yield-factor model* de la estructura a término. Dicho modelo se introduce especificando relaciones simples entre tasas de bonos cupón cero de varios vencimientos fijos. Esto marca una diferencia importante con la mayoría de los modelos multifactorial en que los factores son *observables* desde la curva de rendimientos actual y sus incrementos pueden tener una matriz de correlación especificada arbitrariamente. Este modelo unifica y extiende estrictamente al máximo grado posible modelos afines como los desarrollados en Cox et al. [14], Longstaff and Schwartz [41], entre otros. Además, explota completamente la idea de usar tasas como variables de estado e incluye factores de volatilidad estocástica que son combinaciones lineales específicas de los factores (tasas de bonos). Por último, cabe mencionar que el modelo proporciona una estructura a término de precios de derivados que es consistente con la ausencia de arbitraje y computacionalmente tratable. Si bien no se describe una economía cuyo equilibrio general implique el comportamiento de la estructura a término que aparece en el modelo, éste se puede hacer fácilmente a lo largo de las líneas de CIR (por ejemplo) y agregar algo más a lo ya planteado.

Finalizamos esta sección destacando que esta clase afín de modelos se puede extender permitiendo que los coeficientes (K, H, ρ) sean dependientes del tiempo y para casos con saltos en el proceso de estados X . También se puede resolver analíticamente para la distribución de transición de un proceso de variable de estado afín, y para los precios de opciones en bonos cupón asociados y otros securities, usando métodos de transformadas de Fourier. Los modelos afines, además, se utilizan ampliamente en el análisis de default timing y la valuación relacionada a defaultable bonds.

A.3.2. Modelos de tasa forward

Una alternativa importante para los modelos de tasas short fue la propuesta por Ho y Lee, mencionada en la sección previa, la cual modela la evolución de toda la curva de rendimiento en un marco de árbol binomial. La idea básica se tradujo luego en tiempo continuo por Heath, Jarrow y Morton (HJM), quienes desarrollaron un marco bastante más general para el modelado de la dinámica de las tasas de interés que se adapta a una economía con múltiples factores (ver Heath et al. [31]). Precisamente, eligiendo las tasas forward instantáneas como cantidades fundamentales para modelar, ellos derivaron un marco libre de arbitraje para la evolución estocástica de toda la curva de rendimiento, donde las dinámicas de las tasas forward están completamente especificadas a través de sus estructuras de volatilidad instantáneas. Esta es una diferencia importante con las dinámicas para la tasa short de un modelo de un factor libre de arbitraje, donde la volatilidad de la tasa short no es suficiente para caracterizar el modelo de tasa de interés relevante.

Considerando el marco probabilístico establecido en el Capítulo 1 y $\{W_t\}$ un movimiento Browniano bajo la medida de riesgo neutral, las hipótesis básicas del modelo HJM son:

(HJM.1) Para cada vencimiento fijo $T \leq T^*$, la dinámica de la tasa forward instantánea $f(t, T)$ está dada por:

$$f(t, T) = f(0, T) + \int_0^t \alpha(u, T) du + \int_0^t \sigma(u, T) \cdot dW_u, \quad \forall t \in [0, T]$$

para una función medible Borel $f(0, \cdot) : [0, T^*] \rightarrow \mathbb{R}$, y funciones $\alpha : C \times \Omega \rightarrow \mathbb{R}$ y $\sigma : C \times \Omega \rightarrow \mathbb{R}^d$, con $C = \{(u, t) | 0 \leq u \leq t \leq T^*\}$. Esta condición queda expresada en forma diferencial de la siguiente manera :

$$df(t, T) = \alpha(t, T)dt + \sigma(t, T) \cdot dW_t.$$

(HJM.2) Para cada vencimiento T , $\alpha(\cdot, T)$ y $\sigma(\cdot, T)$ son procesos adaptados tal que

$$\int_0^T |\alpha(u, T)| du + \int_0^T |\sigma(u, T)|^2 du < \infty, \quad a.s.$$

Los coeficientes α y σ se pueden introducir como las funciones de las tasas forward, de forma tal que

$$\alpha(u, T) = \alpha(u, T, f(t, T)), \quad \sigma(u, T) = \sigma(u, T, f(t, T)).$$

Hay una importante relación de consistencia entre α y σ . Bajo condiciones puramente técnicas (ver Duffie [23]), se deberá cumplir:

$$\alpha(t, T) = \sigma(t, T) \cdot \int_t^T \sigma(t, u) du.$$

De esta restricción del drift neutral al riesgo se desprende que basta conocer las tasas forward iniciales $\{f(0, s) : 0 \leq s \leq T\}$ y el proceso σ de la tasa forward para determinar los procesos de precios de todos los bonos y derivados de la tasa de interés.

Apéndice B

Análisis sobre los parámetros λ y Λ en el modelo de Nelson y Siegel

Una curva genérica de Nelson y Siegel tiene la forma

$$f_{NS}(\tau) = \beta_1 + \beta_2 \frac{1 - e^{-\lambda\tau}}{\lambda\tau} + \beta_3 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right), \quad \tau \geq 0$$

siendo $\lambda > 0$. Esta fórmula resulta de integrar la forma funcional dada en la ecuación (2.1). Los parámetros β_1 , β_2 y β_3 se pueden interpretar como tres *factores latentes*. A los términos que acompañan a cada factor se los puede considerar como *cargas*.

B.1. Análisis del parámetro λ

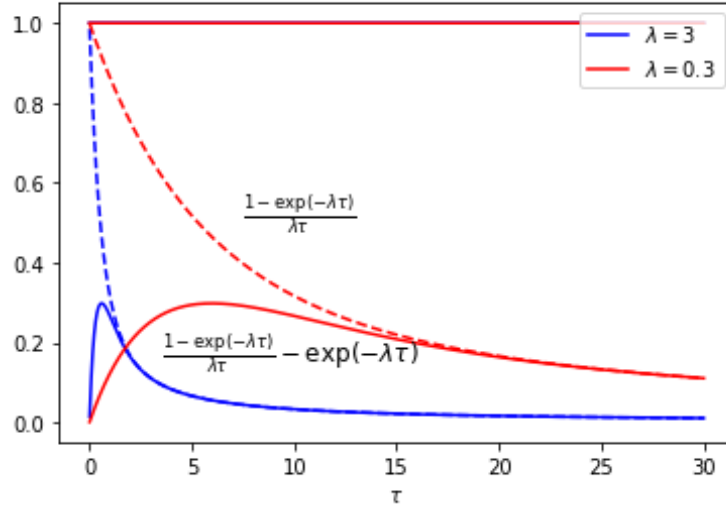
La carga en el primer factor está dada por el valor 1 y las cargas en el segundo y tercer factor, que denominaremos p y c respectivamente, están dadas por

$$p(\tau) = \frac{1 - e^{-\lambda\tau}}{\lambda\tau}, \quad c(\tau) = \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right), \quad \tau \geq 0$$

Nota. La carga p del segundo factor es una función monótona decreciente en τ . La carga del tercer factor posee un único máximo en $\tau_0 = \frac{x_0}{\lambda}$, donde x_0 es la solución positiva de la ecuación

$$x^2 + x + 1 = e^x.$$

El valor de x_0 es aproximadamente $x_0 \sim 1.7933$.

Figura B.1: Gráfico de las cargas para dos valores distintos de λ .

La Figura B.1 muestra el gráfico de las cargas para los 3 factores para distintos valores de λ : $\lambda = 3$ en azul, y $\lambda = 0.3$ en rojo. Es claro que a mayores valores de λ , menor es el punto de máximo de la función c . El valor máximo es siempre el mismo.

Por otra parte, las funciones p y c son ambas decrecientes para $\tau > \tau_0$, y la diferencia entre ellas: $p(\tau) - c(\tau) = e^{-\lambda\tau}$ es positiva y menor a 0.03 para $\tau > 2\tau_0$, con un decrecimiento exponencial. Esto es, ambas funciones tienen el mismo comportamiento asintótico.

Notemos que una escritura equivalente para f_{NS} es

$$f_{NS}(\tau) = \beta_1 + (\beta_2 + \beta_3) \frac{1 - e^{-\lambda\tau}}{\lambda\tau} - \beta_3 e^{-\lambda\tau}, \quad \tau \geq 0.$$

Ajuste de puntos a una curva de Nelson y Siegel

Es frecuente utilizar las curvas de Nelson y Siegel para ajustar una serie de observaciones de rendimientos de bonos $y(\tau_1), y(\tau_2), \dots, y(\tau_n)$, a una curva en particular. Esto implica encontrar valores $\beta_1, \beta_2, \beta_3$ y λ de modo que los errores

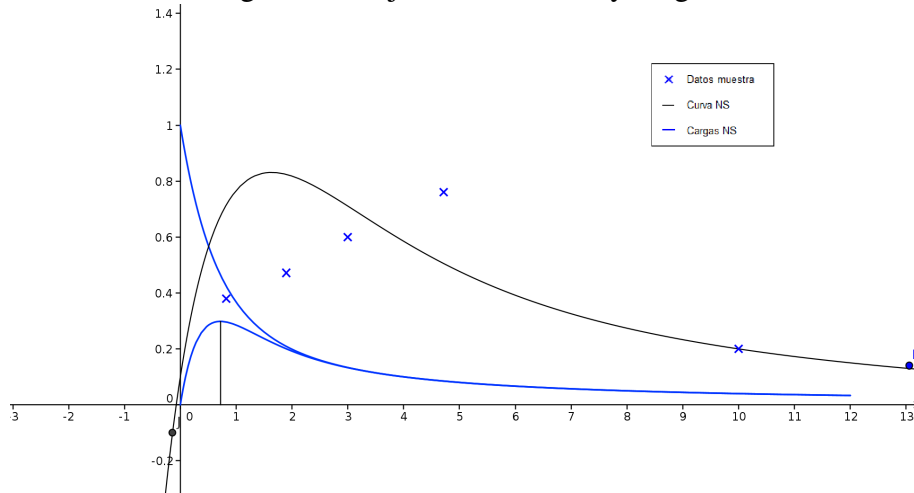
$$|y(\tau_i) - f_{NS}(\tau_i)|, \quad 1 \leq i \leq n,$$

sean mínimos (en algún sentido de mínimo).

Ahora bien, si las observaciones tienen valores máximos o mínimos, es decir, se observa un crecimiento y luego un decrecimiento de los rendimientos (o viceversa), esos cambios no podrán ser capturados por una curva de NS si el λ elegido es *demasiado grande* en relación a estos puntos de cambio, ya que cualquier

curva de Nelson y Siegel tendrá siempre una forma decreciente en el intervalo donde las observaciones tienen valles o picos.

Figura B.2: Ajuste con Nelson y Siegel



Entonces, a fin de modelar estos cambios es conveniente que λ sea elegido de modo que τ_0 esté ubicado en los intervalos de mayores cambios en curvatura de la muestra elegida. Quizás si hay varios máximos y mínimos en una curva en particular convendrá pensar en los puntos donde el cambio porcentual con respecto al punto anterior y siguiente sea más significativo. La Figura B.2 intenta explicar esta situación. El valor de λ se ha elegido de modo que el máximo del tercer factor es sensiblemente menor al punto de máximo (o podría ser de mínimo) del conjunto de observaciones. Independientemente de los valores de β_2 y β_3 que se tomen, no será posible capturar la joroba o valle presente en las observaciones.

Análisis de los extremos de f_{NS}

Analizaremos en esta sección los extremos de la función f_{NS} . Para ello, vemos en qué puntos se anula su derivada.

Tenemos que

$$f'_{NS}(\tau) = (\beta_2 + \beta_3) \frac{e^{-\lambda\tau}(\lambda\tau + 1) - 1}{\lambda\tau^2} + \beta_3 \lambda e^{-\lambda\tau}.$$

Luego $f'_{NS}(\tau) = 0$ si y sólo si

$$(\beta_2 + \beta_3) (e^{-\lambda\tau}(\lambda\tau + 1) - 1) = -\beta_3 (\lambda\tau)^2 e^{-\lambda\tau}.$$

$$(\beta_2 + \beta_3) ((\lambda\tau + 1) - e^{\lambda\tau}) = -\beta_3 (\lambda\tau)^2.$$

$$(\beta_2 + \beta_3) (e^{\lambda\tau} - (\lambda\tau + 1)) = \beta_3 (\lambda\tau)^2.$$

Las expresiones $(e^{\lambda\tau} - (\lambda\tau + 1))$ y $(\lambda\tau)^2$ son no negativas para $\tau > 0$, y ambas valen 0 en $\tau = 0$. Por lo tanto para que exista un extremo de f_{NS} en el intervalo $\tau > 0$ es condición necesaria que β_2 y β_3 tengan el mismo signo.

Notemos que la función

$$g(\tau) = \frac{(e^{\lambda\tau} - (\lambda\tau + 1))}{(\lambda\tau)^2},$$

es siempre positiva, y además creciente. Su derivada es igual a

$$\begin{aligned} g'(\tau) &= \frac{(\lambda e^{\lambda\tau} - \lambda)(\lambda\tau)^2 - (e^{\lambda\tau} - (\lambda\tau + 1))2\lambda^2\tau}{(\lambda\tau)^4} \\ &= \frac{\lambda(e^{\lambda\tau} - 1)\tau - (e^{\lambda\tau} - (\lambda\tau + 1))2}{\lambda^2\tau^3} \\ &= \frac{e^{\lambda\tau-2} - \lambda\tau + 2\lambda\tau + 2}{\lambda^2\tau^3} \\ &= \frac{e^{\lambda\tau}(\lambda\tau - 2) + \lambda\tau + 2}{\lambda^2\tau^3} \\ &> \frac{(\lambda\tau - 2) + (\lambda\tau + 2)}{\lambda^2\tau^3} = \frac{1}{\lambda\tau^2} > 0. \end{aligned}$$

Por otra parte, $\lim_{\tau \rightarrow \infty} g(\tau) = \infty$ y

$$\lim_{\tau \rightarrow 0^+} g(\tau) = \lim_{\tau \rightarrow 0^+} \frac{\lambda e^{\lambda\tau} - \lambda}{2\lambda^2\tau} = \lim_{\tau \rightarrow 0^+} \frac{\lambda e^{\lambda\tau}}{2\lambda} = \frac{1}{2}.$$

Entonces, f'_{NS} tiene un extremo si y sólo si existe un τ_m tal que

$$g(\tau_m) = \frac{\beta_3}{\beta_3 + \beta_2}$$

y esto ocurre si y sólo si

$$\frac{\beta_3}{\beta_3 + \beta_2} \geq \frac{1}{2}.$$

Además, como g es inyectiva, el extremo local si existe es único. Por otra parte, el punto de extremo τ_m es la solución de la ecuación

$$(\lambda\tau)^2 \frac{\beta_3}{\beta_2 + \beta_3} + \lambda\tau + 1 = e^{\lambda\tau}.$$

Teorema B.1.1. Sea f_{NS} la función de Nelson y Siegel dada por

$$f_{NS}(\tau) = \beta_1 + \beta_2 \frac{1 - e^{-\lambda\tau}}{\lambda\tau} + \beta_3 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right), \quad \tau \geq 0$$

con β_1 , β_2 y β_3 constantes, y $\lambda > 0$. Entonces f_{NS} posee un extremo (máximo o mínimo local y absoluto) si y sólo si

$$\frac{\beta_3}{\beta_3 + \beta_2} > \frac{1}{2}. \quad (\text{B.1})$$

En cualquier otro caso, f_{NS} es estrictamente creciente ($\beta_2 < 0$) o estrictamente decreciente ($\beta_2 > 0$).

Si f_{NS} tiene un extremo, entonces ese extremo es

a) Un máximo si $\beta_3 > \beta_2$.

b) Un mínimo si $\beta_3 < \beta_2$.

Demostración. El teorema se sigue del desarrollo previo. Si la derivada no se anula en ningún punto, al ser continua se sigue que es siempre positiva o siempre negativa. Dado que

$$\lim_{\tau \rightarrow 0^+} f_{NS}(\tau) = \beta_1 + \beta_2, \quad \lim_{\tau \rightarrow \infty} f_{NS}(\tau) = \beta_1,$$

se tiene que f es decreciente si $\beta_2 > 0$ y creciente $\beta_2 < 0$.

Para determinar si el extremo es máximo o mínimo, calcularemos la derivada de f_{NS} en 0. Si es positiva, el extremo es un máximo, y si es negativa el extremo es un mínimo.

Tenemos que

$$f'_{NS}(\tau) = (\beta_2 + \beta_3) \frac{e^{-\lambda\tau}(\lambda\tau + 1) - 1}{\lambda\tau^2} + \beta_3\lambda e^{-\lambda\tau}.$$

Aplicando L'Hôpital, la derivada (por derecha) en 0 es igual a:

$$\begin{aligned} f'_{NS}(0) &= \lim_{\tau \rightarrow 0^+} (\beta_2 + \beta_3) \frac{-\lambda e^{-\lambda\tau}(\lambda\tau + 1) + \lambda e^{-\lambda\tau}}{2\lambda\tau} + \beta_3 \lim_{\tau \rightarrow 0^+} (\lambda e^{-\lambda\tau}) \\ &= (\beta_2 + \beta_3) \frac{-\lambda}{2} + \beta_3 \lambda \\ &= \frac{\lambda}{2} (\beta_3 - \beta_2). \end{aligned}$$

- f_{NS} tiene pendiente positiva en $\tau = 0$ si y sólo si $\beta_3 > \beta_2$.
- Para un mismo valor de $\beta_3 - \beta_2$, a mayor valor de λ , mayor es la pendiente. Esto se explica, porque al incrementar λ el extremo estará más a la izquierda pero los valores de la función son los mismos. Luego tiene que aumentar la pendiente para alcanzar el mismo valor en menor tiempo.)

□

Teorema B.1.2. Sea f_{NS} como en el Teorema B.1.1 tal que se cumple la condición (B.1), y sea τ_m el extremo de f_{NS} . Sea x_0 la solución positiva de la ecuación

$$x^2 + x + 1 - e^x = 0.$$

Entonces,

a) Si $\frac{1}{2} < \frac{\beta_3}{\beta_2 + \beta_3} < 1$, entonces $\tau_m < \frac{x_0}{\lambda}$,

b) Si $\frac{\beta_3}{\beta_2 + \beta_3} > 1$, entonces $\tau_m > \frac{x_0}{\lambda}$,

c) Si $\beta_2 = 0$, entonces $\tau_m = \frac{x_0}{\lambda}$.

Demostración. Dado $k \in \mathbb{R}$, denotamos con s_k a la función

$$s_k(x) = kx^2 + x + 1 - e^x = (k-1)x^2 + \underbrace{(x^2 + x + 1 - e^x)}_{u(x)}.$$

Tenemos que $u(x)$ es positiva en el intervalo $[0, x_0)$, negativa para $x > x_0$ y $u(x_0) = 0$. Por lo tanto, si $k > 1$ se tiene que $s_k(x)$ es positiva en $[0, x_0]$. Dado que $\lim_{x \rightarrow \infty} s_k(x) = -\infty$, $s_k(x)$ se anula en algún $x > x_0$.

Por otro lado, si $\frac{1}{2} < k < 1$, entonces $s_k(x)$ es positiva en algún intervalo $(0, \epsilon)$, con $\epsilon > 0$. Esto puede verse analizando el crecimiento de $kx^2 + x + 1$ en relación a e^x . Ambas funciones comienzan en 0, su derivada primera es igual a 1, y la derivada segunda en 0 es $2k$ y 1 respectivamente.

Dado que $s_k(x_0) < 0$, se sigue que se anula en algún punto x , con $0 < x < x_0$.

Con estos resultados tenemos que, si τ_m es tal que

$$(\lambda\tau_m)^2 \frac{\beta_3}{\beta_2 + \beta_3} + \lambda\tau_m + 1 = e^{\lambda\tau_m},$$

entonces $\lambda\tau_m < x_0$ si $\frac{1}{2} < \frac{\beta_3}{\beta_2 + \beta_3} < 1$ y $\lambda\tau_m > x_0$ si $\frac{\beta_3}{\beta_2 + \beta_3} > 1$. □

Corolario B.1.3. La función f_{NS} tiene un extremo si y sólo si $-|\beta_3| < \beta_2 < |\beta_3|$.

Sobre el valor extremo de f_{NS}

Notemos que el extremo de f_{NS} , en caso que exista, ocurre en un punto τ_m que satisface la ecuación

$$k(\lambda\tau)^2 + \lambda\tau + 1 = e^{\lambda\tau},$$

donde

$$k = \frac{\beta_3}{\beta_2 + \beta_3}.$$

Vale decir, que distintas elecciones de β_2 y β_3 dan una misma *ubicación* del máximo.

Si evaluamos f_{NS} en $\tau = \tau_m$ tenemos que:

$$\begin{aligned} f_{NS}(\tau_m) &= \beta_1 + (\beta_2 + \beta_3) \left(\frac{1 - e^{-\lambda\tau_m}}{\lambda\tau_m} \right) - \beta_3 e^{-\lambda\tau_m} \\ &= \beta_1 + (\beta_2 + \beta_3) e^{-\lambda\tau_m} \left(\frac{\beta_3}{\beta_2 + \beta_3} (\lambda\tau_m) + 1 \right) - \beta_3 e^{-\lambda\tau_m} \\ &= \beta_1 + (\beta_2 + \beta_3) e^{-\lambda\tau_m} + \beta_3 (\lambda\tau_m) e^{-\lambda\tau_m}. \end{aligned}$$

Recordemos que para la existencia del extremo es condición necesaria que $\beta_2 + \beta_3$ y β_3 tengan el mismo signo. En particular, estos coeficientes influyen en el valor máximo o mínimo. Es decir, para un mismo valor de λ , cuanto mayor sean estos coeficientes, mayor será el máximo; y análogamente con el mínimo.

B.2. Propiedad de la matriz Λ

En esta sección detallaremos algunas propiedades relevantes de la matriz Λ , vector aleatorio formado por los factores latentes.

Sean $N \in \mathbb{N}_{\geq 3}$ y $0 < \tau_1 < \tau_2 < \dots < \tau_N$ los vencimientos a considerar. De esta manera,

$$\Lambda = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{pmatrix} \in \mathbb{R}^{N \times 3}, \quad \text{con } \lambda > 0. \quad (\text{B.2})$$

Propiedad B.2.1. *Sea Λ la matriz definida en (B.2). Entonces Λ tiene todas sus entradas positivas*

Demostración. Como ya vimos, $0 < \lambda$ y $0 < \tau_1 < \tau_2 < \dots < \tau_N$. Luego, $0 < \lambda\tau_i$ para todo $i = 1, \dots, N$.

Esto implica que

$$\begin{aligned} e^{\lambda\tau_i} > 1 &\iff e^{-\lambda\tau_i} < 1 \\ &\iff 1 - e^{-\lambda\tau_i} > 0 \\ &\iff \frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} > 0 \quad \forall i = 1, \dots, N. \end{aligned}$$

Resta ver que $\frac{1-e^{-\lambda\tau_i}}{\lambda\tau_i} - e^{-\lambda\tau_i} > 0$ para todo $i = 1, \dots, N$. Pero observamos que:

$$\frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} - e^{-\lambda\tau_i} = \frac{1 - e^{-\lambda\tau_i} - \lambda\tau_i e^{-\lambda\tau_i}}{\lambda\tau_i}.$$

Luego,

$$\begin{aligned} \frac{1 - e^{-\lambda\tau_i} - \lambda\tau_i e^{-\lambda\tau_i}}{\lambda\tau_i} > 0 &\iff 1 - e^{-\lambda\tau_i} - \lambda\tau_i e^{-\lambda\tau_i} > 0 \quad (\text{pues } \lambda\tau_i > 0 \forall i) \\ &\iff 1 > e^{-\lambda\tau_i} + \lambda\tau_i e^{-\lambda\tau_i} \\ &\iff 1 > e^{-\lambda\tau_i}(1 + \lambda\tau_i) \\ &\iff e^{\lambda\tau_i} > 1 + \lambda\tau_i \quad \forall i = 1, \dots, N. \end{aligned}$$

Pero para todo $x \in \mathbb{R}$, $e^x = 1 + x + \frac{x^2}{2!} + R_3(x)$, donde $R_3(x) = \frac{e^\zeta x^3}{3!}$ con $\zeta \in (0, x)$ si $x > 0$ o $\zeta \in (x, 0)$ si $x < 0$, es la fórmula de Lagrange para el resto en el desarrollo en serie de Taylor de e^x .

Luego, como $\lambda\tau_i > 0$ esto implica que $R_3(\lambda\tau_i) > 0$ para todo $i = 1, \dots, N$. De esta manera,

$$e^{\lambda\tau_i} = 1 + \lambda\tau_i + \frac{(\lambda\tau_i)^2}{2!} + R_3(\lambda\tau_i) \tag{B.3}$$

$$> 1 + \lambda\tau_i \quad \forall i = 1, \dots, N. \tag{B.4}$$

Con esto concluimos que la matriz Λ tiene todas sus entradas positivas. □

Observación. Las columnas 1 y 2 de la matriz Λ son l.i. También lo son las columnas 1 y 3.

Esto se puede ver fácilmente ya que la columna 1 tiene todos sus elementos iguales, y no ocurre así con las otras columnas. Esto se deduce de manera sencilla observando que las funciones $h_1(x) = \frac{1-e^{-x}}{x}$ y $h_2(x) = e^{-x}$ son estrictamente decrecientes para $x > 0$:

$$h_1'(x) = \frac{e^{-x}(1+x) - 1}{x^2} = \frac{(1+x) - e^x}{e^x x^2}.$$

Pero $e^x > 1 + x$ si $x > 0$ (ver (B.3) en Propiedad B.2.1). Luego, h_1 es decreciente.

Por otro lado,

$$h_2'(x) = -e^{-x} < 0 \quad \forall x > 0.$$

El siguiente Teorema demuestra que la matriz $\Lambda \in \mathbb{R}^{N \times 3}$ tiene rango completo, es decir, es de rango 3.

Teorema B.2.2. *Sea Λ la matriz definida en (B.2). Entonces Λ tiene rango máximo.*

Demostración. Para ver esto, hacemos una simplificación al problema: reemplazamos la tercer columna por $(-1) \times (\text{columna 3} - \text{columna 2})$. Es decir, le aplicamos operaciones elementales por columnas que

no alteran el rango de la matriz. De esta manera, queremos ver que la matriz

$$\begin{aligned} \Lambda_1 &= \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} - e^{-\lambda\tau_N} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} & e^{-\lambda\tau_N} \end{pmatrix} \quad \text{con } \lambda, \tau_i > 0 \text{ para todo } i = 1, \dots, N. \end{aligned}$$

tiene rango completo.

Por la observación previa, las columnas 1 y 2 de la matriz Λ_1 son l.i. También lo son las columnas 1 y 3.

Supongamos ahora que existe una combinación lineal de las tres columnas de Λ_1 igual a 0. Esto significa que existen a, b, c tal que

$$\begin{cases} a + b \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} + ce^{-\lambda\tau_1} = 0 \\ a + b \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} + ce^{-\lambda\tau_2} = 0 \\ \vdots = 0 \\ a + b \frac{1-e^{-\lambda\tau_N}}{\lambda\tau_N} + ce^{-\lambda\tau_N} = 0 \end{cases} \quad (\text{B.5})$$

Luego, el sistema (B.5) es equivalente al sistema:

$$\begin{cases} a(\lambda\tau_1) + b(1 - e^{-\lambda\tau_1}) + c(\lambda\tau_1) e^{-\lambda\tau_1} = 0 \\ a(\lambda\tau_2) + b(1 - e^{-\lambda\tau_2}) + c(\lambda\tau_2) e^{-\lambda\tau_2} = 0 \\ \vdots = 0 \\ a(\lambda\tau_N) + b(1 - e^{-\lambda\tau_N}) + c(\lambda\tau_N) e^{-\lambda\tau_N} = 0 \end{cases} \quad (\text{B.6})$$

Por lo tanto, la matriz Λ o Λ_1 tiene rango máximo si y solo si el sistema (B.6) tiene como única solución $a = b = c = 0$.

Supongamos que existe una solución no trivial. Luego, deben ser c o b no nulos pues si $b = c = 0$, entonces debe ser $a = 0$. Pero por la observación anterior se debe cumplir que $b \neq 0$ y $c \neq 0$.

Consideremos la función $g : \mathbb{R} \rightarrow \mathbb{R}$ continua e infinitamente diferenciable dada por

$$g(x) = ax + b(1 - e^{-x}) + cxe^{-x}.$$

Si existe una solución no trivial de (B.6), entonces la función g se anula en los puntos $x_i = \lambda\tau_i$, para $1 \leq i \leq N$.

Además es fácil ver que $g(0) = 0$:

$$g(0) = a \cdot 0 + b(1 - e^{-0}) + c \cdot 0 \cdot e^{-0} = 0.$$

Por lo tanto, si $N \geq 3$, entonces g se anula en al menos 4 puntos distintos.

Por otro lado, la función g' tiene un único máximo (o mínimo, según como sean b y c), que separa dos intervalos donde g'

$$g'(x) = a + be^{-x} + c(e^{-x}(1 - x)) = a + e^{-x}(b + c(1 - x))$$

es monótona creciente y monótona decreciente, respectivamente. Para ver esto basta analizar la derivada segunda de g :

$$g''(x) = e^{-x}(-b - c(1 - x) + c) = e^{-x}(-b + cx).$$

Vemos que g'' cambia de signo en $x_0 = \frac{b}{c}$. Recordemos que b y c deben ser ambos distintos de 0.

por lo cual a la derecha de x_0 la función g' es estrictamente creciente y a la izquierda estrictamente decreciente, o al revés. Esto significa que g' tiene a lo sumo dos ceros distintos.

Utilizando el mismo razonamiento, concluimos que g tiene a lo sumo 3 ceros distintos. Esto es absurdo puesto que g tiene al menos 4 ceros dados por $x = 0$ y $x = \lambda\tau_i$ para $1 \leq i \leq N$ con $N \geq 3$.

Luego el sistema (B.6) tiene sólo solución trivial y las columnas de Λ_1 son l.i. concluyendo que la matriz Λ_1 y en consecuencia Λ tienen rango 3. \square

Corolario B.2.3. Sean i, j y k tal que $1 \leq i < j < k \leq N$. Luego, las filas i, j y k de Λ son linealmente independientes.

Demostración. Este resultado se obtiene directamente de la demostración anterior observando que la función

$$g(x) = ax + b(1 - e^{-x}) + cxe^{-x},$$

tiene 4 ceros distintos en $0, \lambda\tau_i, \lambda\tau_j$ y $\lambda\tau_k$ si suponemos que el sistema

$$\begin{cases} a(\lambda\tau_i) + b(1 - e^{-\lambda\tau_i}) + c(\lambda\tau_i)e^{-\lambda\tau_i} = 0 \\ a(\lambda\tau_j) + b(1 - e^{-\lambda\tau_j}) + c(\lambda\tau_j)e^{-\lambda\tau_j} = 0 \\ a(\lambda\tau_k) + b(1 - e^{-\lambda\tau_k}) + c(\lambda\tau_k)e^{-\lambda\tau_k} = 0 \end{cases} \quad (\text{B.7})$$

tiene solución no trivial.

Pero vimos en la demostración del lema que esto no es posible pues g tiene a lo sumo tres ceros distintos. Luego, la solución del sistema (B.7) es la trivial y las filas i, j y k de Λ_1 (y en consecuencia de Λ) son linealmente independientes. \square

Apéndice C

Descripción de test estadísticos

En este apéndice detallaremos los test estadísticos que utilizamos en el presente trabajo.

C.1. Test lag óptimo: Criterios de información

El enfoque general es ajustar modelos VAR(m) con órdenes $m = 0, \dots, p_{max}$ y elegir un estimador del orden p que minimice el criterio preferido. Muchos de los criterios de uso actual tienen la forma general

$$Cr(m) = \log \det \tilde{\Sigma}(u(m)) + c_T \phi(m),$$

donde $\det()$ denota el determinante, \log es el logaritmo natural, $\tilde{\Sigma}(u(m)) = T^{-1} \sum_{t=1}^T \hat{u}_t \hat{u}_t'$ es el estimador de la matriz de covarianza residual para un modelo de orden m , c_T es una sucesión que depende del tamaño de muestra T , y $\phi(m)$ es una función que penaliza órdenes grandes del VAR. Por ejemplo, $\phi(m)$ puede representar el número de parámetros que deben estimarse en un modelo VAR(m). El término $\log \det \tilde{\Sigma}(u(m))$ mide el ajuste de un modelo con orden m . Debido a que no hay corrección para los grados de libertad en el estimador de la matriz de covarianza, el determinante del logaritmo decrece (o al menos no aumenta) cuando m aumenta. El tamaño de la muestra debe mantenerse constante; por tanto, el número de valores de muestra previa reservados para la estimación está determinado por el orden máximo p_{max} . Los siguientes criterios son algunos de los más utilizados en la práctica:

$$\begin{aligned} AIC(m) &= \log \det \tilde{\Sigma}(u(m)) + \frac{2}{T} mK^2, \\ HQ(m) &= \log \det \tilde{\Sigma}(u(m)) + 2 \frac{\log \log T}{T} mK^2, \\ SC(m) &= \log \det \tilde{\Sigma}(u(m)) + \frac{\log T}{T} mK^2 \end{aligned}$$

El criterio AIC sobreestima asintóticamente el orden con probabilidad positiva, mientras que los dos últimos criterios estiman el orden de manera consistente bajo condiciones bastante generales si el

proceso real tiene un orden VAR finito y el orden máximo p_{max} es mayor que el orden verdadero. Estos resultados no sólo son válidos para los procesos $I(0)$ sino también para los procesos $I(1)$ con variables cointegradas. Denotando por $\hat{p}(AIC)$, $\hat{p}(HQ)$ y $\hat{p}(SC)$ los órdenes seleccionados por los tres criterios, respectivamente, encontramos que las siguientes relaciones son válidas incluso en muestras pequeñas de tamaño fijo $T \geq 16$ (ver Lütkepohl [43]):

$$\hat{p}(SC) \leq \hat{p}(HQ) \leq \hat{p}(AIC).$$

C.2. Test raíz unitaria: Dickey-Fuller Aumentado

Consideremos un proceso $AR(p)$ y_t tal que

$$y_t = \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + u_t,$$

donde u_t es un proceso de ruido blanco con media cero con la varianza invariante en el tiempo $E(u_t^2) = \sigma_u^2$ y los coeficientes α_i son fijos.

Luego, el proceso es integrado cuando $(1 - \alpha_1 L - \dots - \alpha_p L^p)y_t = u_t$ como vimos en la Sección 3.1.2. En otras palabras, una hipótesis de interés es $\alpha(1) = 1 - \alpha_1 - \dots - \alpha_p = 0$. Para testear esta hipótesis nula contra la alternativa de estacionariedad del proceso, es útil reparametrizar el modelo. Restar y_{t-1} en ambos lados y reorganizar los términos da como resultado una regresión

$$\Delta y_t = \phi y_{t-1} + \sum_{j=1}^{p-1} \alpha_j^* \Delta y_{t-j} + u_t, \quad (C.1)$$

donde $\phi = -\alpha(1)$ y $\alpha_j^* = -(\alpha_{j+1} + \dots + \alpha_p)$. En este modelo deseamos probar el par de hipótesis $H_0 : \phi = 0$ versus $H_1 : \phi < 0$. El *test estadístico Dickey-Fuller aumentado (ADF)* se basa en el estadístico t del coeficiente ϕ de una estimación OLS de (C.1). Más precisamente, el t -ratio de $\hat{\phi}$,

$$TestADF = \frac{\hat{\phi}}{std(\hat{\phi})}$$

donde $\hat{\phi}$ denota la estimación por OLS de ϕ . No tiene una distribución asintótica normal estándar, pero tiene una distribución límite no estándar. Los valores críticos se han obtenido mediante simulación, análisis univariante de series de tiempo y están disponibles, por ejemplo, en Fuller [28] y Davidson and MacKinnon [17]. Sin embargo, resulta que la distribución límite depende de los términos deterministas que deben incluirse. Por lo tanto, se utilizan valores críticos diferentes cuando se agrega un término de tendencia constante o lineal en (C.1). Por otro lado, la inclusión de variables ficticias estacionales además de una tendencia constante o lineal no da lugar a cambios adicionales en la distribución límite.

C.3. Cointegración: Test de Johansen

Siguiendo la sección 8.2 de Lütkepohl [43] y con base en la Proposición 3.4.1, es fácil derivar el estadístico del máxima verosimilitud o Likelihood Ratio (LR) para probar un rango de cointegración específico $r = r_0$ de un VECM contra un rango mayor de cointegración, digamos $r = r_1$. Considere el VECM sin términos deterministas,

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t,$$

donde y_t es un proceso de dimensión K , $rk(\Pi) = r$ con $0 \leq r \leq K$, los Γ_j 's ($j = 1, \dots, p-1$) son matrices de parámetros ($K \times K$) y $u_t \sim N(0, \Sigma_u)$ es ruido blanco Gaussiano, como vimos en el Capítulo 3. Para simplificar, asumimos que el proceso comienza en el tiempo $t = 1$ con valores iniciales cero (es decir, $y_t = 0$ para $t \leq 0$). Alternativamente, los valores iniciales pueden ser valores fijos.

Supongamos que deseamos probar

$$H_0 : rk(\Pi) = r_0 \text{ contra } H_1 : r_0 < rk(\Pi) \leq r_1. \quad (C.2)$$

Bajo supuestos de normalidad, el máximo de la función de verosimilitud para un modelo con rango de cointegración r se da en la Proposición 3.4.1. De ese resultado, el estadístico LR para el test (C.2) viene dado por:

$$\begin{aligned} \lambda_{LR}(r_0, r_1) &= 2[\log l(r_1) - \log l(r_0)] \\ &= T \left[-\sum_{i=1}^{r_1} \log(1 - \lambda_i) + \sum_{i=1}^{r_0} \log(1 - \lambda_i) \right] \\ &= -T \sum_{i=r_0+1}^{r_1} \log(1 - \lambda_i), \end{aligned} \quad (C.3)$$

donde $l(r_i)$ denota el máximo de la función de probabilidad gaussiana para el rango de cointegración r_i . Obviamente, el valor del test es bastante fácil de calcular, utilizando los valores propios de la Proposición 3.4.1. Sin embargo, resulta que la distribución asintótica del estadístico LR bajo la hipótesis nula para r_0 y r_1 dados no es estándar. En particular, no es una distribución χ^2 . Depende del número de tendencias comunes $K - r_0$ bajo H_0 y de la hipótesis alternativa. Dos pares diferentes de hipótesis han recibido la mayor atención en la literatura relacionada:

$$H_0 : rk(\Pi) = r_0 \text{ versus } H_1 : r_0 < rk(\Pi) \leq K \quad (C.4)$$

y

$$H_0 : rk(\Pi) = r_0 \text{ versus } H_1 : rk(\Pi) = r_0 + 1. \quad (C.5)$$

El estadístico LR $\lambda_{LR}(r_0, K)$ para verificar (C.4) a menudo se denomina *estadístico de traza* para testear el rango de cointegración y $\lambda_{LR}(r_0, r_0 + 1)$ se denomina *estadístico de valor propio máximo*. Johansen [35] (1988, 1995) muestra que las distribuciones nulas límites son funcionales de un proceso de Wiener estándar $(K - r_0)$ dimensional. Los puntos porcentuales de las distribuciones asintóticas y, por tanto, los valores críticos para las pruebas LR se pueden generar fácilmente. Las tablas, por ejemplo, están disponibles en Johansen [35].

La estrategia para determinar el rango de cointegración de un sistema dado de K variables es testear una sucesión de hipótesis nulas,

$$H_0 : rk(\Pi) = 0, H_0 : rk(\Pi) = 1, \dots, H_0 : rk(\Pi) = K - 1$$

y terminar los tests cuando la hipótesis nula no pueda rechazarse por primera vez. El rango de cointegración se elige en consecuencia. Aquí se pueden utilizar tanto el valor propio máximo como los tests de trazas. Por ejemplo, si hay tres variables ($K = 3$), primero testamos $rk(\Pi) = 0$. Si esta hipótesis nula no puede rechazarse, el análisis procede con un rango de cointegración de $r = 0$ y, por lo tanto, un modelo en las primeras diferencias se consideran en el análisis posterior. Sin embargo, si se rechaza $rk(\Pi) = 0$, probamos $rk(\Pi) = 1$. Si el test no rechaza esta hipótesis, el análisis puede proceder con un VECM con rango de cointegración $r = 1$. De lo contrario, $rk(\Pi) = 2$ se testea y se elige $r = 2$ como rango de cointegración si esta hipótesis no puede rechazarse. Si también se rechaza $rk(\Pi) = 2$, se puede considerar trabajar con un modelo VAR estacionario para los niveles de las variables. Claramente, en estos tests se debe conocer el orden de rezago del modelo VAR.

C.3.1. Términos deterministas

Comenzamos asumiendo que el término determinista consiste en un término medio constante simple

$$\mu_t = \mu_0.$$

Aunque normalmente pensamos en μ_0 como un vector fijo distinto de cero ($K \times 1$), el caso $\mu_0 = 0$ no se excluye explícitamente. En otras palabras, el usuario del test no está seguro de que la media del proceso sea cero y, por lo tanto, permite la posibilidad de un término medio distinto de cero. En la Sección 3.3.2 está descrito el VECM para este caso para las variables observables y_t . Esto se puede ver en la ecuación (3.24). Por lo tanto, el estadístico LR para testear el rango de cointegración se puede determinar exactamente como en el caso de media cero considerado previamente, excepto que y_{t-1} tiene que ser reemplazado por y_{t-1}^0 en las fórmulas relevantes a partir de las cuales se calculan los valores propios en la Proposición 3.4.1.

Un proceso con una tendencia lineal también es interesante desde un punto de vista práctico. Por tanto, sea $\mu_t = \mu_0 + \mu_1 t$, donde μ_0 y μ_1 son vectores arbitrarios ($K \times 1$). En el Capítulo 3, fórmula (3.28), está desarrollada la expresión del VECM para el proceso observable y_t .

Por lo tanto, los estadísticos de LR de interés se pueden determinar nuevamente exactamente como en el caso de media cero reemplazando y_{t-1} con y_{t-1}^+ y teniendo en cuenta el término de intercepto agregando una fila de unos en ΔX_t en las fórmulas relevantes de la Proposición 3.4.1.

Otra vez, los valores críticos se pueden encontrar en Johansen [35]. Un ejemplo:

Table 15.2 Quantiles of the Likelihood Ratio Test for Cointegration (6.32) With Restricted Constant and $F = (B_1, \dots, B_{p-r}, 1)'$

$p - r$	50.0%	75.0%	80.0%	85.0%	90.0%	95.0%	97.5%	99.0%
1	3.43	5.28	5.85	6.54	7.50	9.13	10.73	12.73
2	11.34	14.45	15.31	16.39	17.79	19.99	22.07	24.74
3	23.15	27.50	28.65	30.05	31.88	34.80	37.56	40.84
4	38.90	44.39	45.86	47.60	49.92	53.42	56.57	60.42
5	58.47	65.19	66.92	69.01	71.66	75.74	79.60	83.93
6	81.85	89.71	91.75	94.13	97.17	101.84	106.07	111.38
7	109.11	118.12	120.52	123.26	126.71	132.00	136.69	142.34
8	140.01	150.10	152.75	155.86	159.74	165.73	171.18	177.42
9	174.65	185.99	188.92	192.27	196.66	203.34	209.28	216.08
10	213.14	225.67	228.81	232.63	237.35	244.56	250.86	258.31
11	255.33	269.01	272.44	276.46	281.63	289.71	296.70	304.89
12	315.69	319.50	323.90	329.51	333.26	338.10	345.77	354.32

Para un mayor desarrollo de estos tests, ver Lütkepohl [43] y Johansen [35].

Apéndice D

Resultados: predicción de 2 pasos

En este apéndice mostraremos los resultados de las simulaciones para los mismos estudios que realizamos previamente pero considerando 2 pasos de predicción. Es decir, se simularán caminos de $T - 2$ pasos y se hará la predicción del paso $T - 1$ y T .

D.1. Caso A: Nivel y MBS

D.1.1. Modelo VAR

Nota. La media estimada corresponde a la media muestral obtenida con los datos de predicción para los valores en $T - 1$ y en T en cada simulación. La std muestral corresponde a la desviación estándar de ese estimador.

Tabla D.1: Resultados del modelo VAR que incluye el factor 1 o nivel y driver MBS con datos diarios desde el 25/08/2011 hasta el 04/08/2014.

Resultados de predicción L_{740}	$tol = 0.03$	$tol = 0.01$	$tol = 0.003$
Lag VAR (BIC)	2	2	2
Número de Simulaciones	100	332	3362
Media Estimada Factor 1 (%)	(-0.1238)	(-0.124)	(-0.1137)
Sesgo Factor 1	(-0.0555)	(-0.0557)	(-0.0455)
Media Estimada Driver MBS (Dur)	0.5038	0.4941	0.5007
Sesgo Driver MBS	0.0238	0.0141	0.0207
Std Muestral Factor 1	0.0174	0.01	0.003
Std Muestral Driver MBS	0.0108	0.0063	0.0019
Resultados de predicción L_{741}	$tol = 0.03$	$tol = 0.01$	$tol = 0.003$
Lag VAR (BIC)	2	2	2
Número de Simulaciones	100	332	3365
Media Estimada Factor 1 (%)	(-0.1234)	(-0.124)	(-0.1136)
Sesgo Factor 1	(-0.0589)	(-0.0595)	(-0.0491)
Media Estimada Driver MBS (Dur)	0.5038	0.4942	0.5007
Sesgo Driver MBS	0.0248	0.0151	0.0216
Std Muestral Factor 1	0.0174	0.01	0.003
Std Muestral Driver MBS	0.0108	0.0063	0.0019

D.1.2. Modelo VECM

Tabla D.2: Resultados del modelo VECM que incluye el factor 1 o nivel y el driver MBS con datos diarios desde el 25/08/2011 hasta el 04/08/2014 asumiendo intercepto y tendencia fuera de la ecuación de cointegración.

Resultados de predicción L_{740}	$tol = 0.03$	$tol = 0.01$	$tol = 0.003$
Lag VECM (BIC)	2	2	2
Número of Simulaciones	100	212	2113
Media Estimada Factor 1 (%)	(-0.045)	(-0.046)	(-0.0397)
Sesgo Factor 1	0.0232	0.0223	0.0285
Media Estimada Driver MBS (Dur)	0.4486	0.4388	0.4606
Sesgo Driver MBS	(-0.0314)	(-0.0412)	(-0.0194)
Std Muestral Factor 1	0.0142	0.008	0.0024
Std Muestral Driver MBS	0.0269	0.018	0.0056
Resultados de predicción L_{741}	$tol = 0.03$	$tol = 0.01$	$tol = 0.003$
Lag VECM (BIC)	2	2	2
Número de Simulaciones	100	211	2109
Media Estimada Factor 1 (%)	(-0.0439)	(-0.0447)	(-0.039)
Sesgo Factor 1	0.0206	0.0198	0.0255
Media Estimada Driver MBS (Dur)	0.449	0.4398	0.4608
Sesgo Driver MBS	(-0.0304)	(-0.0396)	(-0.0185)
Std Muestral Factor 1	0.0142	0.008	0.0024
Std Muestral Driver MBS	0.0269	0.0181	0.0056

D.2. Caso B: Pendiente y USD/JPY

D.2.1. Modelo VAR

Nota. La media estimada corresponde a la media muestral obtenida con los datos de predicción para los valores en $T - 1$ y en T en cada simulación. La std muestral corresponde a la desviación estándar de ese estimador.

Tabla D.3: Resultados del VAR que incluye el factor 2 o pendiente y el driver USD/JPY con data diaria desde 28/03/2013 to 13/08/2015.

Resultados de predicción S_{599}	$tol = 0.03$	$tol = 0.01$	$tol = 0.003$
Lag VAR (BIC)	2	2	2
Número de Simulaciones	100	795	8313
Media Estimada Factor 2 (%)	0.6754	0.647	0.6603
Sesgo Factor 2	0.0534	0.0249	0.0382
Media Estimada Driver USD/JPY	1.2612	1.2513	1.2591
Sesgo Driver USD/JPY	0.02	0.0101	0.0179
Std Muestral Factor 2	0.0297	0.01	0.003
Std Muestral Driver USD/JPY	0.0154	0.0053	0.0016
Resultados de predicción S_{600}	$tol = 0.03$	$tol = 0.01$	$tol = 0.003$
Lag VAR (BIC)	2	2	2
Número de Simulaciones	100	796	8326
Media Estimada Factor 2 (%)	0.6758	0.648	0.661
Sesgo Factor 2	0.0702	0.0423	0.0553
Media Estimada Driver USD/JPY	1.2616	1.2517	1.2594
Sesgo Driver USD/JPY	0.0172	0.0073	0.0151
Std Muestral Factor 2	0.0297	0.01	0.003
Std Muestral Driver USD/JPY	0.0154	0.0053	0.0016

D.2.2. Modelo VECM

Tabla D.4: Resultados de VECM que incluye el factor 2 o pendiente y el driver USD/ JPY con datos diarios desde el 28/03/2013 hasta el 13/08/2015 asumiendo intercepto y tendencia fuera de la ecuación de cointegración.

Resultados de predicción S_{599}	$tol = 0.03$	$tol = 0.01$	$tol = 0.003$
Lag VECM (BIC)	2	2	2
Número de Simulaciones	100	637	6493
Media Estimada Factor 2 (%)	0.6085	0.6281	0.6289
Sesgo Factor 2	(-0.0136)	0.006	0.0068
Media Estimada Driver USD/JPY	1.2371	1.2431	1.2457
Sesgo Driver USD/JPY	(-0.0041)	0.0019	0.0045
Std Muestral Factor 2	0.0244	0.0089	0.0026
Std Muestral Driver USD/JPY	0.0114	0.0047	0.0014
Resultados de predicción S_{600}	$tol = 0.03$	$tol = 0.01$	$tol = 0.003$
Lag VECM (BIC)	2	2	2
Número de Simulaciones	100	637	6503
Media Estimada Factor 2 (%)	0.61	0.6298	0.6306
Sesgo Factor 2	0.0044	0.0241	0.025
Media Estimada Driver USD/JPY	1.238	1.244	1.2465
Sesgo Driver USD/JPY	(-0.0069)	-0.0009	0.0017
Std Muestral Factor 2	0.0244	0.0089	0.0027
Std Muestral Driver USD/JPY	0.0114	0.0047	0.0014

D.3. Caso C: Curvatura y MBS

D.3.1. Modelo VAR

Nota. La media estimada corresponde a la media muestral obtenida con los datos de predicción para los valores en $T - 1$ y T en cada simulación. La std muestral corresponde a la desviación estándar de ese estimador.

Tabla D.5: Resultados del VAR que incluye el factor 3 o curvatura y el driver MBS con datos diarios desde 12/06/2012 a 15/08/2014.

Resultados de predicción C_{549}	$tol = 0.03$	$tol = 0.01$	$tol = 0.003$
Lag VAR (BIC)	2	2	2
Número de Simulaciones	111	1253	13451
Media Estimada Factor 3 (%)	0.4703	0.4872	0.4869
Sesgo Factor 3	0.2166	0.2334	0.2331
Media Estimada Driver MBS (Dur)	0.4753	0.4609	0.4642
Sesgo Driver MBS	0.0133	(-0.0011)	0.0022
Std Muestral Factor 3	0.03	0.01	0.003
Std Muestral Driver MBS	0.0108	0.0034	0.001
Resultados de predicción C_{550}	$tol = 0.03$	$tol = 0.01$	$tol = 0.003$
Lag VAR (BIC)	2	2	2
Número de Simulaciones	107	1228	13206
Media Estimada Factor 3 (%)	0.4654	0.4874	0.486
Sesgo Factor 3	0.1538	0.1758	0.1751
Media Estimada Driver MBS (Dur)	0.4757	0.4603	0.4642
Sesgo Driver MBS	0.0238	0.0083	0.0122
Std Muestral Factor 3	0.03	0.01	0.003
Std Muestral Driver MBS	0.0109	0.0034	0.001

D.3.2. Modelo VECM

Tabla D.6: Resultados del modelo VECM que incluye el factor 3 o curvatura y el driver MBS con datos diarios desde 12/06/2012 al 15/08/2014 asumiendo intercepto y tendencia fuera de la ecuación de cointegración.

Resultados de predicción C_{549}	$tol = 0.03$	$tol = 0.01$	$tol = 0.003$
Lag VECM (BIC)	2	2	2
Número de Simulaciones	100	336	3224
Media Estimada Factor 3 (%)	0.0604	0.0648	0.0658
Sesgo Factor 3	(-0.1933)	(-0.189)	(-0.188)
Media Estimada Driver MBS (Dur)	0.4794	0.4867	0.473
Sesgo Driver MBS	0.0174	0.0247	0.011
Std Muestral Factor 3	0.0173	0.0052	0.0015
Std Muestral Driver MBS	0.0234	0.012	0.0037
Resultados de predicción C_{550}	$tol = 0.03$	$tol = 0.01$	$tol = 0.003$
Lag VECM (BIC)	2	2	2
Número de Simulaciones	100	327	3083
Media Estimada Factor 3 (%)	0.0557	0.0642	0.0635
Sesgo Factor 3	(-0.2559)	(-0.2475)	(-0.2482)
Media Estimada Driver MBS (Dur)	0.4795	0.4871	0.4731
Sesgo Driver MBS	0.0274	0.035	0.021
Std Muestral Factor 3	0.0167	0.0051	0.0014
Std Muestral Driver MBS	0.0234	0.0123	0.0038

Bibliografía

- [1] C. Almeida and J. Vicente. The role of no-arbitrage on forecasting: Lessons from a parametric term structure model. *Journal of Banking & Finance*, 32(12):2695–2705, 2008.
- [2] A. Ang and M. Piazzesi. A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics*, 50(4):745–787, 2003.
- [3] A. Ang, S. Dong, and M. Piazzesi. *No-Arbitrage Taylor Rules*. Manuscript, Columbia University and Stanford University, 2007.
- [4] R. Bibkov and M. Chernov. No-arbitrage macroeconomic determinants of the yield curve. *Journal of Econometrics*, 159:166–182, 2010.
- [5] F. Black and P. Karasinski. Bond and option pricing when short rates are lognormal. *Financial Analysts Journal*, 47:52–59, 1991.
- [6] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–654, 1973.
- [7] D. Brigo and F. Mercurio. *Interest Rate Models-Theory and Practice With Smile, Inflation and Credit*. Springer-Verlag, Berlin Heidelberg, 2001.
- [8] J. Campbell. A defense of traditional hypotheses about the term structure of interest rates. *The Journal of Finance*, 41(4):183–193, 1986.
- [9] J. Campbell and R. Shiller. Yield spreads and interest rate movements: A bird’s eye view. *The Review of Economic Studies*, pages 495–514, 1991.
- [10] R. A. Carmona. *Statistical Analysis of Financial Data in S-Plus*. Springer, Verlag New York, Inc, 2004.
- [11] J. H. E. Christensen, F. X. Diebold, and G. D. Rudebusch. The affine arbitrage-free class of nelson-siegel term structure models. *Journal of Econometrics*, 164:4–20, 2011.

- [12] J. Cochrane and M. Piazzesi. Bond risk premia. *American Economic Review*, 95:138–160, 2005.
- [13] J. Cox, J. Ingersoll, and S. Ross. A re-examination of traditional hypotheses about the term structure of interest rates. *The Journal of Finance*, 36(4):769–799, 1981.
- [14] J. Cox, J. Ingersoll, and S. Ross. A theory of the term structure of interest rate. *Econometrica*, 53:385–408, 1985.
- [15] Q. Dai and K. Singleton. Specification analysis of affine term structure models. *Journal of Finance*, 55:1943–1978, 2000.
- [16] Q. Dai and K. Singleton. Expectation puzzles, time-varying risk premia, and affine models of the term structure. *Journal of Financial Economics*, 63:415–441, 2002.
- [17] R. Davidson and J. MacKinnon. *Estimation and Inference in Econometrics*. Oxford University Press, London, 1993.
- [18] F. X. Diebold and C. Li. Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130:337–364, 2006.
- [19] F. X. Diebold and G. D. Rudebusch. *Yield Curve Modeling and Forecasting: The dynamic Nelson-Siegel Approach*. Princeton University Press, Princeton-New Jersey, 2013.
- [20] F. X. Diebold, G. D. Rudebusch, and S. B. Aruoba. The macroeconomy and the yield curve: A dynamic latent factor approach. *Journal of Econometrics*, 131(1-2):309–338, 2006.
- [21] L. U. Dothan. On the term structure of interest rates. *Journal of Financial Economics*, 6:59–69, 1978.
- [22] G. Duffee. Term premia and interest rate forecasts in affine models. *Journal of Finance*, 57:405–443, 2002.
- [23] D. Duffie. *Dynamic Asset Pricing Theory*. Princeton University Press, Princeton-New Jersey, 2001.
- [24] D. Duffie and R. Kan. A yield-factor model of interest rates. *Mathematical Finance*, 6(4):379–406, 1996.
- [25] R. F. Engle and C. W. J. Granger. Co-integration and error correction: Representation, estimation and testing. *Econometrica*, 55:251–476, 1987.
- [26] E. Fama. Term premiums in bond returns. *The Journal of Finance*, 13:529–546, 1984.

- [27] P. Feldhütter and D. Lando. Decomposing swap spreads. *Journal of Financial Economics*, 88: 375–405, 2008.
- [28] W. A. Fuller. *Introduction to Statistical Time Series*. John Wiley & Sons, New York, 1976.
- [29] C. W. J. Granger. Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 16:121–130, 1981.
- [30] P. R. Halmos. *Measure Theory*. Springer-Verlag, New York-Heidelberg-Berlin, 1974.
- [31] D. Heath, R. Jarrow, and A. Morton. Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica*, 60:77–105, 1992.
- [32] T. Ho and S. Lee. Term structure movements and pricing interest rate contingent claims. *Journal of Finance*, 41:1011–1029, 1986.
- [33] P. Hördahl, O. Tristani, and D. Vestin. A joint econometric model of macroeconomic and term-structure dynamics. *Journal of Econometrics*, 131:405–444, 2006.
- [34] D. Ji and G. Yin. Weak convergence of term structure movements and the connection of prices and interest rates. *Stochastic Anal.*, 11:61–76, 1993.
- [35] S. Johansen. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press Inc., New York, 1995.
- [36] A. Kobor, L. Shi, and I. Zelenko. What determines u.s. swap spreads. *World Bank Working Paper*, 62, 2005.
- [37] J. Lamperti. *Stochastic Processes: A Survey of the Mathematical Theory*. Springer-Verlag, New York, 1977.
- [38] M. Leippold and L. Wu. Asset pricing under the quadratic class. *Journal of Financial and Quantitative Analysis*, 37:271–295, 2002.
- [39] G. Lindgren. *Stationary Stochastic Processes: Theory and Applications*. CRC Press, Taylor & Francis Group, Florida, 2013.
- [40] R. Litterman and J. Sheinkman. Common factors affecting bond returns. *The Journal of Fixed Income*, 1(1):51–61, 1991.
- [41] F. Longstaff and E. Schwartz. Interest rate volatility and the term structure: A two-factor general equilibrium model. *The Journal of Finance*, 47:1259–1282, 1992.

- [42] F. A. Longstaff. Time-varying term premia and traditional hypotheses about the term structure. *The Journal of Finance*, 45(4):1307–1314, 1990.
- [43] H. Lütkepohl. *New Introduction to Multiple Time Series Analysis*. Springer-Verlag Berlin Heidelberg, 2005.
- [44] H. Lütkepohl and M. Krätzig. *Applied Time Series Econometrics*. Cambridge University Press, 2004.
- [45] J. McCulloch. A reexamination of traditional hypotheses about the term structure: a comment. *The Journal of Finance*, 48:779–789, 1993.
- [46] M. Musiela and M. Rutkowski. *Martingale Methods in Financial Modelling*. Springer-Verlag Berlin Heidelberg, 2005.
- [47] C. R. Nelson and A. F. Siegel. Parsimonious modeling of yield curves. *The Journal of Business*, 60: 473–489, 1987.
- [48] B. Øksendal. *Stochastic Differential Equations: An introduction with Applications. 6th Edition*. Springer-Verlag Berlin Heidelberg, 2003.
- [49] J. O. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer-2nd ed., 2005.
- [50] G. D. Rudebusch. Macro-finance models of interest rates and the economy. *Working Paper Series-Federal Reserve Bank of San Francisco*, 2010.
- [51] G. D. Rudebusch and T. Wu. Accounting for a shift in term structure behavior with no-arbitrage and macro-finance models. *Journal of Money, Credit and Banking*, 39:395–422, 2007.
- [52] G. D. Rudebusch and T. Wu. A macro-finance model of the term structure, monetary policy, and the economy. *The Economic Journal*, 118(530):906–926, 2008.
- [53] J. Saúl Gri. Understanding the dynamics of the term structure of interest rates. *Master's Thesis*, 2009.
- [54] S. Seabold and J. Perktold. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.
- [55] K. J. Singleton. *Empirical Dynamic Asset Pricing*. Princeton University Press, Princeton, 2006.
- [56] R. Stigler. *The term structure of interest rates*. In: *Handbook of Monetary Economics, Vol.I, B.M. Friedman and F.H. Hahn*. eds. North-Holland, Amsterdam New York, 1990.

- [57] D. W. Stroock and S. R. S. Varadhan. *Multidimensional Diffusion Processes*. Springer-Verlag Berlin and New York, 1979.
- [58] L. E. Svensson. Estimating and interpreting forward interest rates: Sweden 1992-1994. *IMF Working Paper No 94/114*, 1994.
- [59] R. S. Tsay. *Analysis of Financial Time Series. 2nd Edition*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2005.
- [60] B. Tuckman. *Fixed Income Securities: Tools for Today's Markets. 2nd Edition*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2002.
- [61] O. Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5:177–188, 1977.