

**INSTITUTO TECNOLÓGICO AUTÓNOMO DE
MÉXICO**



Vector Error-Correction Model

**TESIS
QUE PARA OBTENER EL TÍTULO DE
LICENCIADO EN MATEMÁTICAS APLICADAS**

P R E S E N T A

OMAR DÍAZ LANDA

ASESOR: DR. JUAN JOSÉ FERNÁNDEZ DURÁN

MÉXICO, D.F.

2016

Con fundamento en los artículos 21 y 27 de la Ley Federal del Derecho de Autor y como titular de los derechos moral y patrimonial de la obra titulada “ANÁLISIS DE COMPONENTES PRINCIPALES BASADO EN KERNELS”, otorgo de manera gratuita y permanente al Instituto Tecnológico Autónomo de México y a la Biblioteca Raúl Baillères Jr., autorización para que fijen la obra en cualquier medio, incluido el electrónico, y la divulguen entre sus usuarios, profesores, estudiantes o terceras personas, sin que pueda percibir por tal divulgación una contraprestación.

Itzel Zayil Muñoz Fernández de Córdova

Fecha

Firma

Agradecimientos

A mis papás Javier y Yazmín, y a mi hermano Yaíro, por su amor incondicional, por ser mi ejemplo a seguir, y por estar a mi lado en cada paso.

A mi familia, por su inmenso amor, apoyo, confianza y sabiduría.

A Omar por todo su amor y por compartir su vida y proyectos conmigo.

A mis amigos, por su cariño, hermandad y por hacerme reír cada día.

A mis profesores Rubén Hernández Cid, Fernando Esponda, Ernesto Barrios, y mi colega Luis Díaz, por sus enseñanzas y atención.

Índice general

1. Introducción	1
2. Teoría de Cointegración	2
2.0.1. Series de Tiempo	2
2.0.2. Raíz Unitaria: Pruebas	7
2.0.3. Modelos para series de tiempo univariadas	14
3. Métodos kernel	21
4. Análisis de Componentes Principales basado en kernels	22
5. Conclusiones	23
A. Conceptos	24
B. Código	25
Bibliografía	26

Capítulo 1

Introducción

Capítulo 2

Teoría de Cointegración

Teroía de Cointegración

2.0.1. Series de Tiempo

Es necesario familiarizarse con determinados conceptos de uso frecuente en el análisis de series de tiempo. En la actualidad, es más común el uso del enfoque estocástico que el de descomposición de series, debido a que ha demostrado mayor eficacia en la creación de modelos y flexibilidad para extenderse a modelos para varias series simultáneas.

Se dice que es un **proceso estocástico** si se tiene un conjunto índice lineal¹ T y una colección de variables aleatorias definidas en un espacio de probabilidad tales que para cada variable aleatoria le corresponde un sólo elemento del conjunto índice, es decir, $\{Z_t : t \in T\}$. Por lo tanto, una **serie de tiempo** es un proceso estocástico cuyo conjunto índice es el tiempo.

Es importante aclarar que cuando se tiene una serie de tiempo observada, continúa siendo un proceso estocástico en el sentido de que cada observación

¹Sean t_1 y $t_2 \in T$ entonces $t_1 + t_2 \in T$

es una realización de una variable aleatoria; consecuentemente una serie de tiempo observada es una realización de muchas series posibles.

El comportamiento de un proceso estocástico esta descrito completamente por su función de densidad conjunta, sin embargo, conocerla en el caso de series de tiempo es muy complejo y no es posible obtenerla a partir del producto de las funciones de densidad marginales, ya que las variables son altamente dependientes, en realidad, se trata de la misma variable pero medida en diferentes momentos de tiempo. Debido a lo anterior, el análisis de series de tiempo propone conocer los primeros y segundos momentos, es decir, esperanzas, varianzas y covarianzas para poder caracterizar la serie, pues resumen en buena medida a la función de densidad conjunta. Además, nos darán apertura para introducir uno de los conceptos de mayor importancia, el de **estacionariedad**.

Un **proceso estocástico estrictamente estacionario** es un proceso estocástico cuya distribución de densidad conjunta es invariante ante desplazamientos en el tiempo, de manera que si se calculan las medias, varianzas y covarianzas, se obtendrán los mismos resultados sin importar la longitud y momento de tiempo en que se midan, es decir, las medias y varianzas son constantes en el tiempo y las covarianzas dependen únicamente de la distancia entre los dos periodos de tiempos que se calculan. Cuando no es posible conocer la función de densidad conjunta y se utiliza los primeros y segundos momentos para determinar estacionariedad se le conoce como **estacionariedad de segundo orden** o **estacionariedad débil**. En resumen, si Z_t es un proceso estacionario tiene las siguientes características:

$$\text{Media :} \quad E(Z_t) = \mu \quad (2.1)$$

$$\text{Varianza :} \quad \text{Var}(Z_t) = E[(Z_t - \mu)^2] = \gamma_0 \quad (2.2)$$

$$\text{Covarianza} \quad E(Z_t - \mu)E(Z_{t+k} - \mu) = \gamma_k \quad (2.3)$$

El concepto de estacionariedad es de suma importancia, ya que los

modelos econométricos requieren de series estacionarias para poder analizar el comportamiento de la serie en los periodos observados, generalizarlo a otros periodos de tiempo y ajustar modelos para el pronóstico; tal es el caso de los modelos ARIMA de los cuales se hablará más adelante.

Existen diversas pruebas de estacionariedad, pero en la práctica, las más comunes son el **análisis gráfico** y la **función de autocorrelación**. La primera de ellas es poco confiable, ya que consiste, como su nombre lo indica, en graficar la serie y observar que no muestre tendencias, cambios de nivel o que la viariabilidad aumente con el tiempo, esta prueba sirve para darnos una idea del comportamiento de la serie pero se debe considerar como un paso inicial para un análisis más formal como el cálculo de la función de autocorrealción.

Como mencionamos anteriormente, se deben calcular las medias, varianzas y covarianzas para caracterizar, aunque no completamente, a una serie de tiempo, sin embargo, al solo tener una realización del proceso se debe trabajar con las definiciones muestrales:

$$\text{Media muestral:} \quad \bar{Z} = \frac{1}{N} \sum_{t=1}^N Z_t \quad (2.4)$$

$$\text{Varianza muestral:} \quad \hat{\gamma}_0 = \frac{1}{N} \sum_{t=1}^N (Z_t - \bar{Z})^2 \quad (2.5)$$

$$\text{Covarianza muestral:} \quad \hat{\gamma}_k = \frac{1}{N} \sum_{t=1}^{N-k} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z}) \quad k = 0, 1, 2, \dots \quad (2.6)$$

Con estos términos es posible obtener la función de autocorrealción muestral definida como:

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{N-k} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z})}{\sum_{t=1}^N (Z_t - \bar{Z})^2} \quad (2.7)$$

Si la función de autocorrelación tiende rápidamente a cero, nos enfrentamos ante un proceso estacionario; si por el contrario, el decaimiento de la función es paulatino, quiere decir que las observaciones del proceso

muestran una tendencia que las hace depender de otras observaciones en el tiempo, provocando que el nivel de la serie se vea afectado por dicha tendencia y existan correlaciones distintas a cero en periodos de tiempo distantes. Esto último nos hace pensar que la serie no cumple con alguna de las características de un proceso estacionario de segundo orden.

Cuando una serie es no estacionaria, sólo se puede analizar el periodo observado, no es posible hacer uso de los modelos ARMA y cualquier otro tipo de análisis como por ejemplo una regresión, se debe realizar con particular precaución, ya que se podría estar en presencia de un fenómeno que Yule en 1926² identifica como regresión espúrea. Sin embargo, existen herramientas que ayudan a convertir una serie no estacionaria en estacionaria, dichas herramientas son las diferencias estacionarias y diferencias estacionales.

Sea el proceso Z_t tal que al aplicar el **operador diferencia** ∇ se tiene que $\nabla Z_t = Z_t - Z_{t-1}$. De manera que ∇Z_t corresponde al cambio en el valor de la variable Z del periodo $t - 1$ al periodo t , por lo que cuenta con una observación menos que Z_t . Al aplicar el operador la serie obtenida es estacionaria en cuanto a su nivel, pues elimina la tendencia polinomial adaptiva que se encontraba en el proceso.

Si la tendencia polinomial adaptiva es de orden mayor a uno se debe considerar aplicar el operador tantas veces como sea necesario, es decir, se existe presencia de una tendencia polinomial de orden 2, la serie requiere de tomar dos diferencias para estacionarizarla. Además, es importante mencionar el riesgo que existe con la sobrediferenciación pues conlleva a problemas relacionados con la identificación de un modelo para el proceso, incremento en la varianza y pérdida innecesaria de observaciones, pues se

²Yule, G. U., "Why Do We Sometimes Get Nonsense Correlations Between Time Series-
' A Study in Sampling and the nature of Time Series", Journal of the Royal Statisticar
Society, vol. 89,1926, pp.1-64.

pierde una con cada aplicación del operador.

En ocasiones, nos enfrentamos a series que presentan otro tipo de tendencia que se caracteriza por tener un patron de cambio que se repite con una intensidad similar, en determinados periodos de tiempo año con año. Este fenómeno es llamado **estacionalidad** y para controlarlo es necesario en primera instancia identificar el comportamiento periódico de la serie, ya sea que se trate de meses, trimestres, semestres, aos, periodos vacacionales o cualquier otro hecho que provoque que la serie tenga un comportamiento parecido en periodos definidos cada año, posteriormente, se hace uso del **operador diferencia estacional** ∇_E^k donde E es periodo que contiene la tendencia-ciclo y k es el número de veces necesarias que se debe aplicar el operador diferencia para hacer la serie estacionaria.

El operador diferencia estacional esta definido de la siguiente manera:

$$\nabla_E = Z_t - Z_{t-E} \quad (2.8)$$

Al igual que en el operador diferencia estacionaria ∇ se pierden observaciones cada vez que se aplica, en este caso se trata de E observaciones.

Cuando a un proceso es necesario aplicarle el operador diferencia ∇ una vez para hacerlo estacionario, se dice que es un proceso integrado de orden 1 y se denota $Z_t \sim I(1)$. De manera similar, si se requieren 2 diferencias para estacionarizar la serie se trata de un proceso integrado de orden dos y si necesitamos d diferencias es un proceso integrado de orden d , $Z_t \sim I(d)$, finalmente, si z_t es estacionario nos podemos referir a este proceso como un proceso estacionario o como un proceso integrado de orden cero. En general, el orden de integración depende del mínimo número de diferencias que se deben aplicar al proceso para convertirlo en estacionario. Este concepto de procesos estocásticos integrados trae consigo las siguientes propiedades:

Sean Z_t , Y_t y W_t tres procesos estocásticos

1. Si $Z_t \sim I(0)$ y $Y_t \sim I(1)$ entonces $W_t = (Z_t + Y_t) = I(1)$ es decir, una combinación lineal entre un proceso estacionario y uno no estacionario es no estacionario.
2. Si $Z_t \sim I(d)$ entonces $W_t = (a + bZ_t) = I(d)$, donde a y b son constantes. Una combinación lineal de un proceso no estacionario, es no estacionario. Por lo tanto, la combinación lineal de un proceso estacionario continúa siendo estacionario.
3. Si $Z_t \sim I(d_1)$ y $Y_t \sim I(d_2)$ entonces $W_t = (aZ_t + bY_t) = I(d_2)$ ya que $d_1 < d_2$.
4. Si $Z_t \sim I(d)$ y $Y_t \sim I(d)$ entonces $W_t = (aZ_t + bY_t) = I(d^*)$ donde d^* generalmente es igual a d , aunque en ocasiones $d^* < d$.

Como es posible observar, trabajar con varias series que no tienen el mismo orden de integración puede resultar complicado. Sin embargo, en el siguiente capítulo se presentarán técnicas apropiadas para reducir estos riesgos.

2.0.2. Raíz Unitaria: Pruebas

Consideremos el siguiente proceso determinista:

$$Z_t = a_0 + a_1 Z_{t-1} \quad t = \dots, -2, -1, 0, 1, 2, \dots \quad (2.9)$$

que también puede ser expresado a partir del uso del operador L que proviene del inglés *Lag* y se define como $LZ_t = Z_{t-1}$, por lo tanto $Z_t = a_0 + a_1 Z_{t-1}$ ahora es

$$(1 - a_1 L)Z_t = a_0, \quad t = \dots, -2, -1, 0, 1, 2, \dots \quad (2.10)$$

si se despeja a Z_t del proceso tendríamos la siguiente operación, que al desarrollarla se sigue que

$$\begin{aligned}
(1 - a_1 L)^{-1} a_0 &= \sum_{i=0}^{\infty} (a_1 L)^i a_0 \\
&= a_0 \sum_{i=0}^{\infty} (a_1 L)^i \quad L^i 1 = 1 \\
&= a_0 \sum_{i=0}^{\infty} (a_1)^i \\
&= \frac{a_0}{1 - a_1} \quad |a_1| < 1
\end{aligned} \tag{2.11}$$

En la expresión anterior, el supuesto $|a_1| < 1$ permite la convergencia de la serie geométrica real, sin embargo, a_1 puede tomar cualquier valor excepto $a_1 = 1$. Por lo tanto, la solución general para este proceso sería de la forma

$$Z_t = \frac{a_0}{1 - a_1} + s a_1^t \tag{2.12}$$

donde el término $s a_1^t$ se agrega para poder brindar soluciones particulares a partir de la constante s . Ahora bien, si por un lado conservamos el supuesto de la expresión (11), es decir, que la raíz de la ecuación $(1 - a_1 L) = 0$ cumpla que $|a_1| < 1$ podemos observar que

$$\lim_{t \rightarrow \infty} Z_t = \frac{a_0}{1 - a_1} \tag{2.13}$$

lo cual indica que el proceso determinista alcanzará un punto de equilibrio en el largo plazo. Por otro lado, si $a_1 = 1$ entonces el proceso diverge, ya que el cociente $\frac{a_0}{1 - a_1}$ quedaría indeterminado y si $|a_1| > 1$ el proceso no converge y por lo tanto jamás tenderá a estabilizarse.

Ahora que se entiende la problemática que existe cuando se tienen raíces unitarias en procesos deterministas, es oportuno llevar estos resultados a procesos estocásticos, de manera que si definimos a la constante $a_1 = 1$, la constante $a_0 = (1 - L)Z_0$ y agregamos una variable aleatoria que depende del tiempo e_t^3 con media cero, varianza constante σ_e^2 y no correlacionada,

³ Mejor conocida como ruido blanco

tendríamos al siguiente proceso:

$$Z_t = Z_{t-1} + a_0 + e_t \quad (2.14)$$

que a su vez puede ser reexpresada como

$$\begin{aligned} (1 - L)Z_t &= a_0 + e_t \\ (1 - L)Z_t &= (1 - L)Z_0 + e_t \\ (1 - L)\tilde{Z}_t &= e_t \quad \text{con } \tilde{Z}_t = Z_t - Z_0 \end{aligned} \quad (2.15)$$

Este proceso contiene el valor obtenido en el periodo pasado mas un error en cada una de las realizaciones, de manera que cada realización de la variable aleatoria e_t en cada momento t proporciona una tendencia estocástica que provoca que el proceso tenga fluctuaciones, por lo tanto, no es correcto decir que el proceso alcance un punto de equilibrio, sino que se tendrá que recurrir al concepto de estacionariedad ya conocido. En este particular caso, las fluctuaciones implicarían que si tomamos medias muestrales en diferentes intervalos de tiempo, obtendríamos diferentes resultados, característica que nos indica que la serie es no estacionaria. Para comprobar lo anteriormente mencionado, calculamos los dos primeros momentos del proceso después de haber inferido el comportamiento de la serie con el método iterativo.

$$\begin{aligned} E(\tilde{Z}_t) &= E(e_1 + e_2 + \dots + e_t) = 0 \\ Var(\tilde{Z}_t) &= Var(e_1 + e_2 + \dots + e_t) = t\sigma_e^2 \end{aligned} \quad (2.16)$$

Por lo que podemos observar que no cumple con los supuestos que requiere la estacionariedad de segundo orden, pues la varianza aumenta conforme se

avanza en el tiempo causando que el proceso no regrese a su nivel esperado. Por el contrario, si el valor de la constante a_1 es menor que la unidad, entonces el proceso es estacionario y se demuestra calculando nuevamente los momentos de primer y segundo orden, para facilitar la operación reexpresamos al proceso usando la serie geometrica de la siguiente manera:

$$\tilde{Z}_t = (1 - a_1 L)^{-1} e_t = e_t + a_1 e_{t-1} + a_1^2 e_{t-2} + a_1^3 e_{t-3} + \dots \quad (2.17)$$

Por lo tanto, la media y la varianza del proceo son

$$\begin{aligned} E(\tilde{Z}_t) &= E(e_t) + a_1 E(e_{t-1}) + a_1^2 E(e_{t-2}) + \dots = 0 \\ Var(\tilde{Z}_t) &= \sigma_e^2 (1 + a_1^2 + a_1^4 + \dots) = \frac{\sigma_e^2}{(1 - a_1^2)} \end{aligned} \quad (2.18)$$

Que demuestran estacionariedad de segundo orden ya que ninguna depende del tiempo. Por lo anterior, es de nuestro interés hacer pruebas sobre el valor de a_1 para determinar si es estadísticamente igual a la unidad, por lo tanto, en la prueba de hipotiesis correspondiente, se tendría como hipótesis nula el hecho de que $a_1 = 1$ que implicaría no estacionariedad en el proceso, mientras que la hipótesis alternativa abarca los posibles valores de a_1 que proporcionan un proceso estable:

$$H_0 : a_1 = 1 \quad H_a : |a_1| < 1 \quad (2.19)$$

No obstante es posible reparametrizar al proceso sustrayendo en cada lado de la ecuación el valor Z_{t-1} , con la finalidad de que la prueba sea similar a la prueba sobre coeficientes individuales de regresión

$$\begin{aligned}
Z_t - Z_{t-1} &= a_1 Z_{t-1} - Z_{t-1} + e_t \\
\nabla Z_t &= (a_1 - 1) Z_{t-1} + e_t \\
\nabla Z_t &= \rho Z_{t-1} + e_t
\end{aligned} \tag{2.20}$$

donde $\rho = (a_1 - 1)$ de manera que la nueva hipótesis nula y alternativa son:

$$H_0 : \rho = 0 \quad H_a : |\rho| < 1 \tag{2.21}$$

Así que $\rho = 0$ implica que $a_1 = 1$ tratándose de un proceso no estacionario, mientras que si se rechaza la hipótesis nula tendremos un proceso estable, la parte $|\rho| > 1$ no se considera pues la serie sería explosiva. Esta última prueba es la llamada **Prueba de Dickey-Fuller** y es complemento formal a los métodos ya mencionados como la gráfica de los valores de la serie respecto al tiempo, el uso del autocorrelograma y la varianza muestral para determinar estacionariedad, así como el número de diferencias necesarias para convertir la serie en estacionaria.

Para hacer uso de la prueba, es imprescindible conocer el valor del coeficiente ρ pero como se dijo anteriormente, la prueba es en esencia la misma que para los coeficientes de una regresión, así que se estima este valor por medio de mínimos cuadrados. Si $\rho > 0$ implica que $a_1 > 1$ por lo que se puede concluir rápidamente que la serie es no estacionaria, no obstante, si $\rho \leq 0$ se tiene que observar si es estadísticamente diferente de cero, así que se calcula el estadístico de prueba

$$\tau_0 = \frac{\hat{\rho}}{se(\hat{\rho})} \tag{2.22}$$

Si el proceso es estacionario, se puede continuar con el análisis usual de las pruebas sobre coeficientes individuales de una regresión, sin embargo, cuando el proceso no lo es, se tiene la particularidad de que no se distribuye como una $t - Student$, ya que si se aceptara la hipótesis nula, Z_t sería un proceso no estacionario cuya varianza aumenta conforme se incrementa el tamaño de muestra, provocando que la distribución usual $t - Student$ se modifique. La contribución de Dickey y Fuller consistió en la creación de los valores críticos del estadístico τ para tres diferentes procesos a partir de simulaciones de Monte Carlo. Dichos procesos corresponden en primer lugar, al mostrado en la ecuación (19).

En segundo lugar, al proceso

$$\nabla Z_t = \alpha + \rho Z_{t-1} + e_t \quad (2.23)$$

donde cada realización de la variable ∇Z_t depende del valor de una constante, más una proporción del valor anterior Z_{t-1} , más el error e_t . Estos procesos comunmente presentan tendencias definidas hacia arriba si el valor de α es positivo y tendencias definidas decrecientes si es negativo, es decir, tienen una tendencia determinística, ya que se le agrega un valor constante α en cada instante de tiempo t . En este caso, si rechazamos la hipótesis nula, es posible hablar de un proceso estacionario pero con media distinta de cero.

Finalmente, tenemos al proceso que incluye constante y una tendencia en el tiempo

$$\nabla Z_t = \alpha + \beta t + \rho Z_{t-1} + e_t \quad (2.24)$$

como es posible observar, el nuevo término proporciona impacto a la tendencia en cada instante de tiempo, si la prueba de hipótesis refleja que el coeficiente $\rho \neq 0$ entonces se trata de un proceso estacionario al rededor de una tendencia determinística. Para los tres casos se usa la misma prueba de hipótesis y el mismo estadístico de prueba pero diferentes valores críticos que

dependen del tamaño de la muestra, tal y como se muestra en la siguiente tabla. El estadístico τ debe ser más negativo que el valor crítico correspondiente para rechazar la hipótesis nula ($\tau \leq \tau^*$).

Tabla 2.1: Valores Críticos para la prueba Dickey-Fuller

RoyalBlueModel	1 %	5 %	10 %
Ceruleanheight $\nabla Z_t = \rho Z_{t-1} + e_t$	-2.56	-1.94	-1.62
$\nabla Z_t = \alpha + \rho Z_{t-1} + e_t$	-3.43	-2.86	-2.57
$\nabla Z_t = \alpha + \beta_1 t + \rho Z_{t-1} + e_t$	-3.96	-3.41	-3.13
Ceruleanheight <i>Standard critical values</i>	-2.33	-1.65	-1.28
Ceruleanheight			

Tiempo después, los supuestos de la prueba de Dickey-Fuller se vieron relajados permitiendo que existiera correlación entre los errores, la cual puede surgir cuando el modelo no captura la dinámica completa del proceso, es decir, le faltan elementos para explicar a la variable. Lo anterior se corrige agregando un mayor número de diferencias de primer orden, tantas como sean necesarias para asegurar que la correlación en los residuos desaparezca, estas se pueden determinar con el gráfico de las autocorrelaciones considerando todas aquellas que sean grandes y descartando de manera formal las que no son necesarias con la significancia de los coeficientes β_i , obteniendo así el siguiente proceso general:

$$\nabla Z_t = \alpha + \beta_1 t + \rho Z_{t-1} + \sum_{i=1}^m \beta_i \nabla Z_{t-i} + e_t \quad (2.25)$$

Analogamente a la prueba no aumentada, se debe procurar no incluir o excluir el término constante y a la tendencia determinística cuando no es necesario, ya que podría traer problemas de sesgo al estadístico τ y pérdida de significancia de la prueba si se eligen erróneamente los valores críticos, es posible hacer uso de los mismos valores mostrados en el cuadro 2.0.2 **num tabla**

, ya que la prueba de Dickey-Fuller aumentada sigue la misma distribución asintótica.

2.0.3. Modelos para series de tiempo univariadas

La metodología propuesta por Box y Jenkins se convirtió en una herramienta muy popular para el análisis de series de tiempo, cuya principal función era el pronóstico de valores futuros y modelar los valores observados de una serie de tiempo. Estos modelos son conocidos como **autoregresivos e integrados de promedios móviles**, ya que están formados por una parte autoregresiva considerando p retrasos, una parte de q choques aleatorios en el tiempo y, si es necesario, las d diferencias que indique el orden de integración para que el proceso sea estacionario en cuanto a su nivel; de manera que lo anterior usualmente es referido como $ARIMA(p, d, q)$.

Modelos Autoregresivos

Como su nombre lo indica, se trata de una regresión lineal con la diferencia de que las variables independientes, en este caso, son retrasos de la variable dependiente, es decir, la variable se explica a partir de la historia de sus valores pasados. El caso más simple, es el que encontramos en la ecuación (2.0.2) **num equ 14** denominado como autoregresivo de orden 1 $AR(1)$, ya que sólo es explicado por una proporción de su valor retrasado un periodo más el efecto que brinda el ruido blanco.

$$\tilde{Z}_t = \phi \tilde{Z}_{t-1} + e_t \quad (2.26)$$

Es necesario recordar que para tener una serie estacionaria el valor de ϕ debe ser menor a la unidad en valor absoluto y que la media y la varianza están dados por:

$$E(\tilde{Z}_t) = 0 \quad (2.27)$$

$$Var(\tilde{Z}_t) = \frac{\sigma_e^2}{1 - \phi^2} \quad (2.28)$$

Así que la varianza de la serie de tiempo autoregresiva de orden uno depende del valor de la constante ϕ y de la varianza de e_t ; para el cálculo de las autocovarianzas entre dos series Z que se encuentran separadas por un periodo tenemos que

$$\begin{aligned} Cov(\tilde{Z}_t, \tilde{Z}_{t-1}) &= E[\tilde{Z}_t \tilde{Z}_{t-1}] \\ &= E[(e_t + \phi e_{t-1} + \phi^2 e_{t-2} + \phi^3 e_{t-3} + \dots)(e_{t-1} + \phi e_{t-2} + \phi^2 e_{t-3} + \phi^3 e_{t-4} + \dots)] \\ &= \phi E(e_{t-1}^2) + \phi^3 E(e_{t-2}^2) + \phi^5 E(e_{t-3}^2) + \dots \\ &= \phi \sigma_e^2 (1 + \phi^2 + \phi^4 + \dots) \\ Cov(\tilde{Z}_t, \tilde{Z}_{t-1}) &= \frac{\phi \sigma_e^2}{1 - \phi^2} \end{aligned} \quad (2.29)$$

por lo tanto, el autocorrelograma obtiene sus valores mediante la división de las autocovarianzas entre la varianza de la serie de tiempo, esto es

$$\rho_k = \frac{Cov(\tilde{Z}_t, \tilde{Z}_{t-k})}{Var(\tilde{Z}_t)} = \phi^{|k|} \quad k = \pm 0, \pm 1, \pm 2, \dots \quad (2.30)$$

de manera que la correlación entre errores que se encuentran k periodos separados, es el valor de ϕ elevado a la k -ésima potencia. El autocorrelograma tenderá a mostrar comportamientos diferentes dependiendo de los valores que tomen las raíces del polinomio característico de la serie de tiempo, es decir, si las raíces son positivas y menores que uno, el decaimiento será exponencial, por el contrario, si son negativas y mayores a -1 el decaimiento será alternadamente. El panorama deseado consiste en que el decaimiento del autocorrelograma tienda a cero de manera exponencial, lo

cual se verá favorecido mientras se incrementa el valor de k , sin embargo, en ocasiones algunas correlaciones resultan ser significativamente diferentes de cero, lo que nos lleva a pensar que nuestro modelo autoregresivo de orden 1 no captura en buena medida la dinámica del proceso y por lo tanto es necesario introducir una mayor cantidad de retrasos, por ejemplo p . A esto se le conoce como modelo autoregresivo de orden p , $AR(p)$, así que el orden del modelo es igual al retraso más largo de la serie Z .

$$\tilde{Z}_t = \theta_1 \tilde{Z}_{t-1} + \theta_2 \tilde{Z}_{t-2} + \theta_3 \tilde{Z}_{t-3} + \cdots + \theta_p \tilde{Z}_{t-p} + e_t \quad (2.31)$$

al igual que en el caso anterior, es posible determinar la función de autocorrelación en función de los coeficientes autoregresivos del modelo $AR(p)$, que deberá mostrar un decaimiento exponencial a cero cuando las raíces del polinomio característico asociado a la serie se encuentren fuera del círculo unitario, es decir, es estacionario.

Una característica que poseen todos los modelos autorregresivos conocida como invertibilidad consiste en poder ser representados, como se puede observar en la ecuación num tabla 17, como una suma infinita ponderada de choques aleatorios:

$$\tilde{Z}_t = e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \psi_3 e_{t-3} + \cdots \quad \text{con} \quad \sum_{i=1}^{\infty} |\psi_i| < \infty \quad (2.32)$$

Esta expresión explica exactamente lo mismo que la representación con parámetros finitos, por lo que es recomendable trabajar con ésta última. Sin embargo, se da apertura para pensar que existan fenómenos económicos, financieros o incluso naturales que puedan ser modelados a partir de choques aleatorios ponderados, los cuales son conocidos en el análisis de series de tiempo como **Modelos de Promedios Móviles**.

Modelos Promedios Móviles

Es posible hallar otros mecanismos que pudieron haber generado al proceso estocástico Z_t como por ejemplo, a partir de una combinación lineal de los términos del error a lo largo del tiempo, es decir, como una suma finita ponderada de choques aleatorios independientes.

$$\tilde{Z}_t = \theta(L)e_t = (1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_q L^q)e_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} \quad (2.33)$$

Donde \tilde{Z}_t son las desviaciones de Z_t respecto a su nivel medio μ y θ_i para $i = 1, 2, \dots, q$ son las ponderaciones asociadas a los choques aleatorios independientes. Esto quiere decir que dado un proceso que se encuentra en equilibrio, presenta fluctuaciones causadas por los choques aleatorios los cuales no necesariamente deben volver al nivel de la serie en el periodo inmediato posterior sino que esto sucederá únicamente en el largo plazo, además la intensidad reflejada en cada uno de los términos del ruido blanco es causada por el valor del coeficiente θ_i correspondiente. Es natural el querer asociar dichas fluctuaciones independientes con eventos fortuitos que provocan que la serie se aleje de su nivel medio aunque sea momentáneamente. La función que ocupa el término de ruido blanco e_t difiere en cada uno de los modelos que hemos visto hasta ahora, ya que un solo impacto de choque aleatorio en el modelo de promedios móviles $MA(q)$ afecta al proceso estocástico Z en el periodo actual y a lo más en q periodos futuros, mientras que en un modelo autoregresivo $AR(p)$ influye infinitamente en los valores futuros, pues el término e_t se encuentra en el proceso Z_t, Z_{t+1}, Z_{t+2} y así sucesivamente.

Es menester presentar una de las características de mayor importancia del modelo de promedios móviles, para ello debemos recordar la representación de los modelos $AR(p)$ como suma infinita ponderada de choques aleatorios de la ecuación 17, la cual tiene como restricción que $\sum_{i=1}^{\infty} |\psi_i| < \infty$ para poder conocer los primeros momentos del proceso, sin embargo, todo modelo MA

satisface dicha condición y si calculamos la media, varianza y covarianzas del modelo $MA(1)$ podemos observar lo siguiente:

El proceso $MA(1)$ es de la forma

$$\tilde{Z}_t = (1 - \theta_1 L)e_t = e_t + \theta_1 e_{t-1} \quad (2.34)$$

A partir del cual se puede conocer su media y su varianza de manera inmediata

$$E(\tilde{Z}_t) = E(e_t + \theta_1 e_{t-1}) = E(e_t) + \theta_1 E(e_{t-1}) = 0 \quad (2.35)$$

$$\begin{aligned} Var(\tilde{Z}_t) &= Var(e_t - \theta_1 e_{t-1}) \\ &= Var(e_t) + Var(\theta_1 e_{t-1}) \\ &= \sigma_e^2 + \theta_1^2 \sigma_e^2 \\ &= \sigma_e^2 (1 + \theta_1^2) \end{aligned} \quad (2.36)$$

Para el caso de las autocovarianzas se tiene que

$$E[\tilde{Z}_t \tilde{Z}_{t-k}] = E[(e_t - \theta_1 e_{t-1})(e_{t-k} - \theta_1 e_{t-k-1})] = \begin{cases} -\theta_1 \sigma_e^2 & \text{si } k = 1 \\ 0 & \text{si } k \geq 2 \end{cases} \quad (2.37)$$

Por lo que debemos notar que ninguno de sus momentos depende del tiempo, resultado que se puede extender fácilmente a modelos de promedios móviles de ordenes mayores a 1. Por lo tanto, es posible identificar como característica principal de los Modelos MA que todos son estacionarios.

Finalmente, la función de autocorrelación es de la forma

$$\rho_k = \begin{cases} -\frac{\theta_1}{1+\theta_1^2} & \text{si } k = 1 \\ 0 & \text{si } k \geq 2 \end{cases} \quad (2.38)$$

Al observar la *FAC* podemos observar que sólo la primera autocorrelación es diferente de cero y las siguientes son todas iguales a cero, de manera que cumplen con el decaimiento exponencial rápido de procesos estacionarios, además indica que el proceso solo recuerda un periodo anterior, es decir, no tiene memoria. Un valor cercano a uno en la primera autocorrelación puede implicar una fuerte dependencia entre el periodo actual y el periodo anterior, por lo que sería conveniente proponer un modelo autoregresivo ya que podrá captar en mejor medida la dinámica del proceso, por consiguiente es deseable tener una primera autocorrelación pequeña. De igual manera, si se tiene un modelo $MA(q)$ se tendrán las primeras q autocorrelaciones diferentes de cero y el resto iguales a cero, donde las primeras q deberán ser cortas.

Modelos Autoregresivos e Integrados de Promedios Móviles

En la práctica la mayoría de las veces nos enfrentamos con series que no son estacionarias, es decir, que sus primeros momentos no son invariantes en el tiempo. Ya hemos visto que cuando se tienen procesos no estacionarios en cuanto a su nivel es ocasionado, en el algunos casos, por la presencia de una tendencia polinomial adaptiva la cual es posible eliminar mediante la aplicación del operador diferencia un número adecuado de veces, sin embargo, cuando la no estacionaridad es provocada por varianza no constante se debe pensar, como en la regresión lineal múltiple, en una transformación estabilizadora de varianza. Además, usualmente los procesos presentan características tanto de modelos *AR* como de *MA* permitiendo así una generalización que mezcla a los dos modelos anteriores con el concepto de integración.

El modelo $ARIMA(p, d, q)$ es el siguiente:

$$\phi(L)\nabla^d\tilde{Z}_t = \theta(L)e_t \quad d \geq 1 \quad (2.39)$$

donde $\phi(L)$ representa un polinomio de retraso de orden p con parámetros autoregresivos, $\theta(L)$ un polinomio de retraso de orden q con parámetros de promedios móviles y \tilde{Z}_t un proceso con orden de integración d . Es necesaria esta transformación en la variable \tilde{Z}_t ya que los modelos *AR* y

MA sólo pueden operar con procesos estacionarios, ya que de lo contrario no sería posible captar de manera correcta la dinámica del proceso y mucho menos pronosticar para volores futuros. Por lo tanto, al aplicar las d diferencias a un proceso con orden de integración d se convierte estacionario en cuanto a su nivel permitiendo ser respresentado por una parte autoregresiva y una parte de promedios móviles. El motivo principal de incluir a los dos modelos consiste en utilizar el menor número de parámetros para describir el comportamiento de la serie considerando tanto los valores del proceso en el pasado como las fluctuaciones provocadas por choques aleatorios.

El modelo $ARIMA(p, d, q)$ no hereda de manera inmediata las características principales de cada una de sus partes, es decir, no hereda la invertibilidad ni estacionariedad del modelo AR y MA respectivamente. Sin embargo, es factible obtener dichas características siempre y cuando las raíces de los polinomios $\phi(L)$ y $\theta(L)$ se encuentren fuera del círculo unitario.

Capítulo 3

Métodos kernel

Capítulo 4

Análisis de Componentes Principales basado en kernels

Capítulo 5

Conclusiones

Apéndice A

Conceptos

Apéndice B

Código

Bibliografía

- [1] C. M. BISHOP, *Pattern Recognition and Machine Learning*, Springer, 2006.
- [2] N. CRISTIANINI AND J. SHAWE-TAYLOR, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, Cambridge University Press, 2004.
- [3] N. CRISTIANINI AND J. SHAWE-TAYLOR, *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.
- [4] A. GRETTON, *Introduction to RKHS, and Some Simple Kernel Algorithms*. 2015.
- [5] T. HASTIE, R. TIBSHIRANI, AND J. FRIEDMAN, *The Elements of Statistical Learning*, Springer, 2008.
- [6] I. T. JOLLIFE, *Principal Component Analysis*, Springer, 2002.
- [7] M. LICHMAN, *UCI Machine Learning Repository*. <http://archive.ics.uci.edu/ml> , University of California, Irvine, School of Information and Computer Sciences, 2013.
- [8] R. J. E. MERRY, *Wavelet Theory and Applications*. Eindhoven University of Technology, 2005.

- [9] J. NOCEDAL AND S. J. WRIGHT, *Numerical Optimization*, Springer, 2006.
- [10] W. RUDIN, *Functional Analysis*, McGraw-Hill, 1991.
- [11] B. SCHÖLKOPF, A. SMOLA, AND K.-R. MÜLLER, *Kernel Principal Component Analysis*, in *Advances in Kernel Methods: Support Vector Learning*, MIT Press, 1999, pp. 327–352.
- [12] D. SEJDINOVIC AND A. GRETTON, *Foundations of Reproducing Kernel Hilbert Spaces II. Advanced Topics in Machine Learning*. Gatsby Unit, 2012.
- [13] R. S. SHAH, *Support Vector Machines for Classification and Regression*, Master's thesis, McGill University, 2007.
- [14] N. H. TIMM, *Applied Multivariate Analysis*, Springer, 2002.