LA COINTEGRACIÓN EN SERIES DE TIEMPO, UNA APLICACIÓN A LA RELACIÓN ENTRE EL PIB Y EL NIVEL DE EXPORTACIONES EN COLOMBIA

Rios S. Omar A.*

Universidad del Valle, Facultad de Ingeniería, Escuela de Estadística

e-mail: *omar.rios@correounivalle.edu.co*

Abstract: This paper presents the application of model error correction (VEC), aimed at measuring the cointegration of two or more time series, or prove that the characteristics of the series are significantly similar and that shocks in one of them alters to a similar extent the behavior of the other. For the analysis is taken as an example trajectories cointegration for gross domestic product (GDP) and the level of traditional exports in Colombia, including for quarterly periods between January 2000 and the third quarter of 2013. It finds a vector cointegration explaining a common evolution for these two variables in the long run, which concluded that an increase of 1 % in exports generates domestic product growth of 0.22 %.

Keywords: Temporal Causality, Cointegration, Exports, Export Led Growth, GDP

Resumen: El presente documento presenta la aplicación de modelo de corrección de errores (VEC), cuyo objetivo es la medición de la cointegración de dos o más series de tiempo, es decir probar que las características de las series son significativamente similares y que choques en una de ellas altera en una medida similar el comportamiento de la otra. Para el análisis se tomará como ejemplo las trayectorias de cointegración para el producto interno bruto (PIB) y el nivel de exportaciones tradicionales en Colombia, comprendidos para periodos trimestrales entre enero de 2000 y el tercer trimestre de 2013. Se encuentra que existe un vector de cointegración que explica una evolución común para estas dos variables en el largo plazo, donde se concluye que un crecimiento de 1 % en las exportaciones, genera un crecimiento en el producto interno del 0.22 %.

Palabras claves: Causalidad Temporal, Cointegración, Exportaciones, Export Led Growth, PIB

1. INTRODUCCIÓN

El análisis de cointegración es una técnica estadística de amplio uso, sobre todo en ciencias económicas ya que resuelve el problema de la medición de las relaciones de equilibrio entre las variables en el largo plazo tomando en cuenta las características no estacionarias de las mismas. También resuelve el problema estadístico de las regresiones espurias, el cual ocurre cuando se trabaja con variables que se encuentran integradas en algún nivel.

El principal objetivo de este análisis, es realizar una revisión teórica del tema con las distintas propiedades estadísticas del mismo y desarrollar una aplicación práctica con dos variables económicas que teóricamente presenten una relación en el largo plazo.

En economía se hace relevante la relación de las variables en niveles, la cointegración plantea la forma estructural de los componentes que mantienen las variables articuladas en el tiempo, en otras palabras estima la relación entre los comportamientos similares de las variables sin eliminar la información producida al trabajar las variables en diferencias.

Se establece que dos o más variables están cointegradas, si estas se mueven de manera conjunta en el tiempo y la diferencia entre ellas es estable o estacionaria, aun cuando estas contengan un comportamiento tendencial estocástico y por ende no estacionario.

La cointegración se trabaja a partir de un modelo de corrección de errores (VEC), que tiene como objetivo identificar y estimar las relaciones de equilibrio en el largo plazo, de dos o más series de tiempo que teóricamente presenten un comportamiento similar.

Como ejemplo de aplicación para este análisis, se estimará un vector de cointegración que relaciona el crecimiento del PIB a precios constantes con el nivel de las EXPORTACIONES tradicionales en Colombia para el periodo comprendido entre los años 2000 y 2013 con temporalidad trimestral. En la teoría neoclásica se denomina*"Export Led Growth"*, a la correlación positiva entre estas dos variables [Alonso and Patiño, 2007].

Se supone que teóricamente estas dos variables presentan una relación de equilibrio en el largo plazo, debido a que en los últimos años se ha establecido una asociación de causalidad, resultado del diseño de políticas de libre comercio, encaminadas al crecimiento económico [Rendón, 2007].

El presente documento consta de cuatro partes, la primera, esta parte introductoria donde se explica a grandes rasgos el objetivo de la cointegración y la elección de las variables para el desarrollo del método estudiado, en la segunda parte se describe de teóricamente los principales componentes del modelo de cointegración como lo son las pruebas de raíz unitaria y el modelo de corrección de errores, en la tercera parte se realiza el ejemplo aplicado con las variables ya referenciadas en el paquete R utilizando la librería vars [Pfaff, 2008] y urca [Pfaff and Stigler, 2013], interpretando cada resultado, la cuarta parte se realizan las conclusiones del modelo de cointegración y su aplicación.

* Profesor Escuela de Estadística, Universidad del Valle, Colombia. Estadístico, Ms. Economía

2. METODOLOGÍA

Para estimar la posible relación de equilibrio en el largo plazo entre dos o más variables medidas en el tiempo estas deben tener el mismo orden de integración, es decir el número de veces que necesitan ser diferenciadas para ser un proceso estacionario debe ser el mismo, para comprobar este supuesto se recurre a las pruebas de raíz unitaria. La siguiente base metodológica se basa en [Pfaff, 2006], [Enders, 2004] y [Mauricio, 2007].

2.1 *Series de tiempo no estacionarias y pruebas de raíz unitaria*

La mayor parte de las variables macroeconómicas, son procesos no estacionarios, lo cual implica que no tienen media ni varianza constantes en el tiempo y la correlación entre dos observaciones distintas depende del tiempo y no sólo de la distancia (mismo número de períodos).

El grado de integración de una serie de tiempo se refiere al número de diferencias que se deben realizar, para que la variable se convierta en un proceso estacionario.

Supongamos que se tiene una serie de tiempo Y_t como una realización de un componente determinístico y un componente estocástico.

$$Y_t = T_t + x_t \tag{1}$$

Donde $T_t = \beta_1 + \beta_2 t$ es el componente que representa una tendencia determinística y x_t corresponde a una estructura estocástica por ejemplo del tipo $ARMA(p,q)$:

$$\Phi(L)x_t = \Phi(L)\varepsilon_t \varepsilon_t \sim iid \tag{2}$$

Si las raíces que resuelven el polinomio característico del anterior proceso se encuentran por fuera del círculo unitario, es decir mayores que el valor absoluto de la unidad, entonces se concluye que $\{Y_t\}$ es un proceso estacionario alrededor de una tendencia determinística. Si se remueve la tendencia y se realiza la estimación sólo para el proceso ARMA se tendría que $\{Y_t\}$ es un proceso estacionario del tipo $I(0)$.

Ahora si se asume que una de las raíces del proceso $AR(p)$ se encuentra en el círculo unitario y que las restantes $(p-1)$ se encuentran por fuera de manera que:

$$\Delta x_t = (1-L)x_t \Rightarrow \phi(L)\Delta x_t = \theta(L)\varepsilon_t\varepsilon_t \sim iid$$

Entonces Δx_t es un proceso estacionario alrededor de una media constante y el proceso $\{Y_t\}$ es estacionario en diferencia. En otras palabras $\{Y_t\}$ es un proceso con una raíz unitaria del tipo $I(1)$.

Un ejemplo de un proceso $I(1)$, es el de caminata aleatoria, cuyo estudio es importante para demostrar varias cuestiones importantes como decidir el orden de integración de una serie y determinar si las series de un proceso multivariante se encuentran cointegradas debido a esta componente.

Una caminata aleatoria se define como:

$$Y_t = Y_(t-1) + \varepsilon_t \tag{3}$$

$$\Delta Y_t = \varepsilon_t\varepsilon_t \sim iid(0, \sigma_\varepsilon^2) \tag{4}$$

Así que $Y_t \sim I(1)$, y la solución de 4 viene dada por la siguiente expresión:

$$Y_t = Y_0 + \Sigma_{i=1}^{t}\varepsilon_i = Y_0 + e_t \tag{5}$$

donde e_t es la acumulación de los choques pasados hasta el momento t y $E(Y_t) = Y_0$ y $Var(Y_t) = t\sigma_\varepsilon^2$, de lo cual se deduce que es una caminata aleatoria en un proceso no estacionario, ya que la varianza se encuentra en función del tiempo.

Retomando la ecuación 1 inicialmente planteada:

$$Y_t = T_t + x_t$$

$$x_t = T_t + Y_t$$

Ahora si se elimina la tendencia y se asume que x_t sigue un proceso $AR(1)$ se puede describir este proceso como:

$$x_t = \phi x_{t-1} + \varepsilon_t \tag{6}$$

Note que para $\phi = 1, \{x_t\}$ representa una caminata aleatoria con orden de integración $I(1)$, retomando:

$$x_t - x_{t-1} = \phi x_{t-1} - x_{t-1} + \varepsilon_t \tag{7}$$

Tabla 1. Valores críticos de la distribución **DF**

Distribución	**Nivel de significancia** 10 %	5 %	1 %
τ	-1.61	-1.95	-2.61
τ intercepto	-3.51	-2.89	-2.58
τ int y tendencia	-3.15	-3.45	-4.04
t-student	-1.28	-1.65	-2.33

$$\Delta x_t = (\phi - 1)x_{t-1} + \varepsilon_t \tag{8}$$

$$\Delta x_t = \pi x_{t-1} + \varepsilon_t \tag{9}$$

De lo anterior se deduce que: si $\pi = 0$ se tendría que $\phi = 1$, lo cual corresponde a la hipótesis nula de que el proceso presenta orden de integración $I(1)$, es decir tiene raíz unitaria, la hipótesis alterna corresponde a una serie estacionaria en tendencia si $\phi < 1$ y en consecuencia $\pi < 0$.

El caso donde $\pi > 0$, representa un proceso $\{x_t\}$ de carácter explosivo. El coeficiente π, puede ser estimado de forma común por medio del método de mínimos cuadrados ordinarios.

2.1.1 *Prueba de raíz unitaria Dickey-Fuller (DF)*

Formalmente se puede contrastar lo siguiente, si se tiene que x_t es un proceso $AR(1)$ de la forma:

$$x_t = \phi x_{t-1} + \varepsilon_t$$

Entonces $H_0 : x_t \sim I(1)$ contra $H_a : x_t \sim I(0)$, lo cual es equivalente a $H_0 : \phi = 1$ contra $H_a : \phi < 1$.

Ahora de la ecuación:

$$\Delta x_t = \pi x_{t-1} + \varepsilon_t \tag{10}$$

Probar que $H_0 : x_t \sim I(1)$ contra $H_a : x_t \sim I(0)$, es equivalente a $H_0 : \pi = 0$contra $H_a : \pi < 0$.

Para cualquier proceso $AR(1)$, como el expresado en la ecuación (13), el coeficiente π es calculado por MCO y sigue usualmente una distribución t student,

Para el contraste de raíz unitaria esto no es cierto. En este caso la distribución de π es conocida y recibe el nombre de distribución **Dickey – Fuller (DF)**, el cual se representa con la letra τ (tau).

El contraste **(DF)** rechaza la Ho de existencia de una raíz unitaria, cuando el valor calculado τ es menor al valor crítico de la distribución al nivel de significancia indicado. A continuación se presentan algunos valores críticos calculados por convergencia:

A continuación se presenta una simulación para

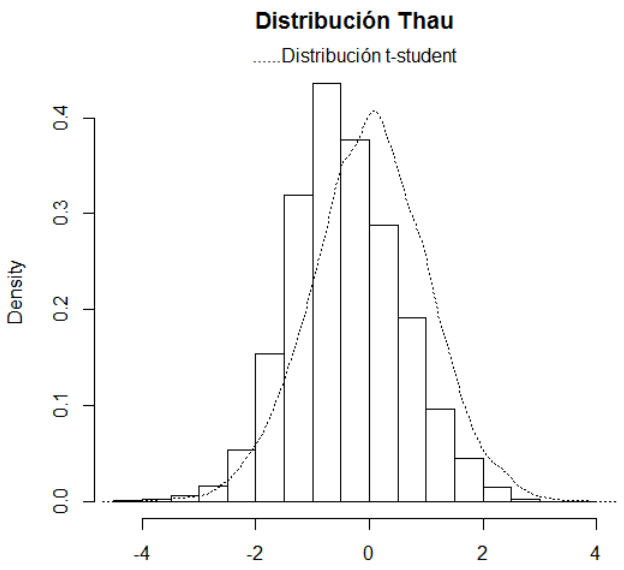

Figura 1. Simulación de la distribución del estadístico Thau y la distribución t-student

establecer un contraste entre la distribución t-student y la distribución τ de la prueba Dickey-Fuller, en donde se puede observar el sesgo de esta última respecto al origen.

Es decir, mientras la distribución t-student es simétrica respecto al origen, la distribución τ es simétrica pero respecto a un valor menor a cero.

2.1.2 *Prueba de raíz unitaria Dickey-Fuller Aumentada (ADF)*

Una extensión del modelo propuesto en la ecuación 10 consiste en permitir un intercepto o "drift", expresado de la siguiente forma:

$$\Delta x_t = \mu + \pi x_{t-1} + \varepsilon_t$$

Donde los valores críticos que permiten el contraste de raíz unitaria se encuentran en la segunda fila de la tabla 1.

Otra modificación a la ecuación 10 es en la que el proceso x_t puede seguir un proceso $AR(p)$, $p \geq 2$:

$$x_t = \mu + \Sigma_{i=1}^{p} \phi_i \Delta x_{t-i} + \varepsilon_t \tag{11}$$

donde:

$$\Delta x_t = \mu + \pi x_{t-}) + \Sigma_{i=1}^{p-} \phi_i^* \Delta x_{t-i} + \varepsilon_t \tag{12}$$

De lo anterior se deduce que $\pi = \Sigma_{i=1}^{p} \phi_i - 1$ y $\phi_i^* = -\Sigma_(j = i+1)^p \phi_j$, entonces si en $\pi = -(1 - \Sigma_{i=1}^{p} \phi_i)$, la $\Sigma \phi_i = 1$ entonces $\pi = 0$, y el proceso presenta evidencia de raíz unitaria.

El contraste se expresa como, $H_0 : x_t \sim I(1)$ contra $H_a : x_t \sim I(0)$. Esta es la prueba **Dickey-Fuller Aumentada (ADF)** y se puede evaluar cuando $\mu = 0$ o $\mu \neq 0$.

En la prueba ADF también se puede implementar una tendencia determinista de la forma:

$$\Delta x_t = \mu + \beta t + \pi x_(t-1) + \Sigma_{i=2}^{p} \beta_i \Delta x_{t-i+1} + \varepsilon_t \tag{13}$$

En la cual asume que Δx_t tiene una tendencia lineal, donde los valores críticos de la prueba se encuentran en la fila 3 de la tabla 1.

Se debe tener cuidado con la escogencia del número de rezagos aplicados en la prueba ADF, ya que de escogerse muy pocos se corre el riesgo que la distribución de los errores del proceso Δx_t no se distribuyan ruido blanco, y la inclusión de un número de rezagos innecesario reduce el poder de la prueba de detectar una raíz unitaria.

2.2 *Cointegración y Modelo de corrección de error (VEC)*

[Engle and Granger, 1987], introducen el concepto de cointegración el cual es un proceso entre dos variables no estacionarias en el tiempo, que existe cuando la combinación lineal entre ellas es estacionaria dado que ambas presentan un mismo orden de integración.

Es decir, si se toma el caso más simple donde se tiene un vector $Z_t = [X_t, Y_t]'$ con $X_t \sim I(1)$ y $Y_t \sim I(1)$ tal qué con la combinación lineal:

$$X_t = \alpha_0 + \alpha_1 Y_t + \hat{\mu}_t$$

Se tiene que, $\hat{u}_t \sim I(0)$ entonces X_t y Y_t se encuentran cointegradas, es decir presentan una relación estable y de equilibrio en el largo plazo.

Los $\hat{u}_t$ representan las perturbaciones respecto a la senda de equilibrio en el largo plazo y α_1 es la elasticidad del largo plazo correspondiente a la relación de equilibrio.

La Cointegración es una propiedad dominante tal que si $x_t \sim I(1)$ y $y_t \sim I(0)$ entonces la combinación lineal de las series es $I(1)$:

$$z_t = x_t + y_t \sim I(1)$$

De manera general si $x_t \sim I(d)$ y $y_t \sim I(e)$ entonces $(x_t + y_t) \sim I(max((d, e))$ si $d \neq e$.

De una manera formal, se dice que dos series x_t y y_t se encuentran cointegradas de orden $d, b, Coin(d, b)$

donde $z_t = (x_t, y_t)$, si:

a) Todos los componentes de z_t son $I(d)$, $x_t \sim I(d)$ y $y_t \sim I(d)$

b) Existe un vector $\alpha(\neq 0)$ tal que: $\alpha_1 x_t + \alpha_2$ y $y_t \sim I(d-b), d > b > 0$, donde α es llamado vector de cointegración.

El principal interés de este procedimiento es el de encontrar relaciones de equilibrio de series no estacionarias en el largo plazo. Si se considera el caso en el que $d = 1$, $b = 1$, $z_t = (x_t, y_t) \sim Coin(1,1) \Rightarrow x_t \sim I(1), y_t \sim I(1)$

Si existe un vector $\alpha = (\alpha_1, \alpha_2)$ tal que:

$$\alpha_1 x_t + \alpha_2 y_t \sim I(0)$$

entonces el resultado de la combinación lineal es estacionario, aunque las series individuales sean no estacionarias, y se puede concluir que las series x_t y y_t están cointegradas es decir presentan una relación de equilibrio estable del largo plazo.

Una forma eficiente de estimar relaciones de cointegración es la propuesta por [Engle and Granger, 1987]:

En primer lugar, se debe demostrar que $z_t = (x_t, y_t) \sim Coin(1,1)$ a partir de las pruebas de raíz unitaria.

En la segunda etapa, por medio de MCO, calcular la regresión lineal:

$$x_t = \beta_0 + \beta_1 y_t + \varepsilon_t$$

El caso en que $\hat{\varepsilon}_t \sim I(0)$ (no presente raíz unitaria) entonces se concluye que existe cointegración entre x_t y y_t.

Donde el vector de cointegración queda expresado como $\alpha = (1, -\beta_1)$[1], tal que β_1, representa la elasticidad de largo plazo correspondiente a la relación de equilibrio.

En tercer lugar se debe estimar el modelo de corrección de error o modelo de cointegración de Johansen (VEC o ECM). Se debe modelar el corto y largo plazo en la misma dinámica.

Si las variables se encuentran cointegradas (la hipótesis nula de no cointegración es rechazada), los residuales de la regresión de equilibrio se usan estimar el modelo de corrección de error. Si $(x_t, y_t) \sim Coin(1,1)$ las relaciones del corto y largo plazo quedan configuradas de la siguiente forma:

$$\begin{bmatrix} \Delta y_t \\ \Delta x_t \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \begin{bmatrix} 1 - \beta \end{bmatrix} \begin{bmatrix} l y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{21} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \Delta x_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \quad (14)$$

Se observa que en el largo plazo las series se encuentran en sus niveles mientras que en el corto plazo se encuentran en diferencias.

Teóricamente, este procedimiento puede ser usado para medir relaciones de cointegración entre dos o más variables por ejemplo: $Z_t = [Z_{t1}, Z_{t2}, Z_{t3}, \ldots, Z_{tN}]$, y la regresión por MCO será:

$$Z_{t1} = \beta_0 + \beta_2' Z_{t2} + \epsilon_t$$

Donde $Z_{t2} = [Z_{t2}, Z_{t3}, \ldots, Z_{tN}]'$, $\beta_2' = [\beta_2, \ldots, \beta_N]'$ y el vector de cointegración es $\alpha = [1, -\beta_2']'$.

En general la estimación del VEC es:

$$\Delta Z_t = \alpha\beta' Z_{t-1} + \Gamma_1 \Delta Z_{t-1} + \ldots + \Gamma_p \Delta Z_{t-p} + \varepsilon_t \quad (15)$$

Para determinar si existe o no cointegración del anterior vector autoregresivo, se verifica si el rango de la matriz Π corresponde al rango dado por la matriz $\alpha\beta'$:

$$Y(\Pi) = Y(\alpha\beta')$$

Para explicar lo anterior nos debemos apoyar en una definición más amplia de la cointegración, "Un vector de variables Z_t de tamaño $(n \times 1)$, se encuentra cointegrado, si al menos un elemento diferente de cero que pertenece al vector de cointegración β_i es tal que $\beta_i' Z_t$ es una tendencia estacionaria. Si existen r vectores tal que $\beta_i (i = 1, 2, \ldots, r)$son linealmente independientes, se dice entonces que $\{Z_t\}$ es cointegrada de rango r. Se define entonces la matriz-$(n \times r)$ de vectores de cointegración $\beta = (\beta_1, \ldots, \beta_r)$. Donde los r elementos del vector $\beta_i' Z_t$ tienen tendencia estacionaria y β es la matriz de cointegración".

Considere el siguiente vector autoregresivo (VAR) de orden K:

$$Z_t = \mu + \Pi_1 Z_{t-1} + \ldots + \Pi_K Z_{t-K} + \Phi D_t + \varepsilon_t \quad (16)$$

Donde Z_t es un vector $(n \times 1)$ de series hasta el periodo t, las matrices $\Pi_i (i = 1, \ldots, K)$ de dimensión $(n \times n)$ son matrices de coeficientes de las variables endógenas rezagadas, μ es un vector de constantes $(n \times 1)$ y D_t es un vector de variables no estocásticas. El término de error $(n \times 1)$ se asume $iid \varepsilon_t \sim N(0, \Sigma)$.

[1] Suponga que x_t y y_t están cointegradas $(1,1)$ $\exists\, \alpha_1, \alpha_2 : \alpha_1 x_t + \alpha_2 y_t \sim I(0)$, entonces dado que α_1 y α_2 no son únicos se debe normalizar el vector para tener una solución de forma que: $x_t + \frac{\alpha_2}{\alpha_1} y_t$ $I(0)$ o en su defecto $\frac{\alpha_1}{\alpha_(2)} x_t + y_t \sim I(0)$

De la ecuación 16 se desprende la especificación del modelo VEC:

$$Z_t = \mu + \Pi Z_{t-1} + \Gamma_1 \Delta Z_{t-1} + \ldots + \Gamma_{K-1}\Delta Z_{t-K+1} + \Phi D_t + \varepsilon_t \quad (17)$$

dónde:

$\Gamma_i = -(\Pi_{i+1} - \ldots - \Pi_K)$ para $i = 1, \ldots, K-1$,
$\Pi = -(I - \Pi_1 - \ldots - \Pi_K)$

Γ_i Incorpora en el modelo VEC los efectos transitorios o del corto plazo.

Se asume que los componentes individuales de Z_t son al menos $I(1)$, sin embargo el resultado del VEC es estacionario. Se supone que las diferencias presentan comportamiento estacionario y el término de corrección de error ΠZ_{t-1}, debe ser estacionario $I(0)$, de otra forma el VEC no sería equilibrado.

Se debe definir entonces las condiciones de Π para que Z_t sea estacionario. Se consideran para este efecto tres casos:

1. $Y(\Pi) = n$
2. $Y(\Pi) = 0$
3. $0 < Y(\Pi) = r < n$

Donde $Y()$ representa el rango de la matriz. En el primer caso donde el rango de Π es n, todas las combinaciones lineales pueden ser estacionarias y las desviaciones de Z_t alrededor del componente determinístico debe ser estacionario, entonces la ecuaciones (19) representa un vector autoregresivo estándar (VAR) en los niveles de Z_t.

En el segundo caso donde el rango de Π es cero, no existe una combinación lineal tal que ΠZ_t sea estacionaria, excepto por la solución trivial en donde se obtiene un modelo VAR en primeras diferencias.

El tercer caso en el cual $0 < Y(\Pi) = r < n$, que es una matriz de rango incompleto, existen dos matrices α y β' de dimensión $(n \times r)$ tal que $\Pi = \alpha\beta'$. Por lo tanto $\alpha\beta' Z_{t-K}$ es estacionario y el producto de $\beta' Z_{t-K}$ también es estacionario.

Las r columnas linealmente independientes del vector β son los vectores de cointegración y el rango de Π es igual al rango de cointegración del sistema Z_t. Cada columna de la matriz representa las relaciones del largo plazo entre las series individuales de Z_t.

En la práctica se puede obtener el número de vectores de cointegración confirmando la significancia de las raíces características de la matriz Π, ya que el rango de una matriz corresponde al número de raíces características diferentes de cero. Suponga que de la matriz Π se obtienen n raíces características ordenadas de forma $\lambda_1 > \lambda_2 > \ldots > \lambda_n$, si las variables en Z_t no se encuentran cointegradas, el rango de Π es igual a cero, y todas las raíces características serán cero también.

Dado que $ln(1) = 0$, cada expresión $ln(1 - \lambda_i)$ es cero si las variables no están cointegradas. De la misma forma si el rango de la matriz es uno, entonces $0 < \lambda_1 < 1$ y $ln(1 - \lambda_1) < 0$, y para todos los demás $\lambda_i = 0$ y $ln(1 - \lambda_2) = ln(1 - \lambda_3) = \cdots = ln(1 - \lambda_n) = 0$. Las pruebas estadísticas para determinar cuántas raíces características son diferentes de cero se basan en este procedimiento:

La prueba de la traza:

$$\lambda_{tr}(r) = -T\Sigma_{i=r+1}^{n} ln(1 - \hat{\lambda}_i) \quad (18)$$

La prueba del máximo valor propio:

$$\lambda_{ei}(r, r+1) = -T ln(1 - \hat{\lambda}_{r+1}) \quad (19)$$

Dónde:

$\hat{\lambda}_i$ Corresponde a los valores propios estimados de las raíces características de la matriz Π, T es el número de observaciones.

2.3 *Elementos determinísticos en el modelo de cointegración*

Existen cinco posibles combinaciones de componentes determinísticos que pueden ser incluidos en un modelo VEC.

1. Sin componentes determinísticos:

   $$\Delta Z_t = \alpha\beta' Z_{t-1} + \Gamma_1 \Delta Z_{t-1} + \ldots + \Gamma_p \Delta Z_{t-p} + \varepsilon_t \quad (20)$$

   Las series se encuentran distribuidas alrededor de cero, no es común encontrar series económicas que cumplan con este comportamiento.

2. Intercepto en los vectores de cointegración y no componentes determinísticos en el modelo de corto plazo VAR.

   $$\Delta Z_t = \mu_{2t} + \alpha\beta' Z_{t-1} + \Gamma_1 \Delta Z_{t-1} + \ldots + \Gamma_p \Delta Z_{t-p} + \varepsilon_t \quad (21)$$

   Las series fluctúan alrededor de una constante μ_2. La interpretación es que la relación en el largo plazo entre las variables fluctúa alrededor de una constante. μ_2 Corresponde al término que representa los datos en niveles en la regresión.

3. Intercepto en el vector de cointegración e intercepto en VAR.

   $$\Delta Z_t = \mu_1 + \mu_{2t} + \alpha\beta' Z_{t-1)} + \Gamma_1 \Delta Z_{t-1} + \ldots + \Gamma_p \Delta Z_{t-p} + \varepsilon_t \quad (22)$$

   Estas series fluctúan alrededor de una constante, es decir la relación del largo plazo entre las

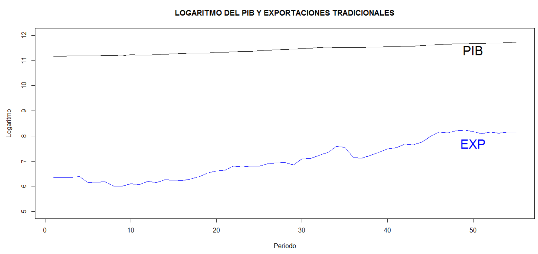

Figura 2. Logaritmo del PIB y logaritmo de las Exportaciones tradicionales

variables que fluctúa alrededor de la constante también presenta tendencia en las series en niveles.

4. Intercepto y tendencia en el vector de cointegración e intercepto en el VAR

$$\Delta Z_t = \mu_1 + \mu_{2t} + \alpha\beta' \binom{Z_{t-1}}{t-1} + \Gamma_1 \Delta Z_{t-1} + \ldots + \Gamma_p \Delta Z_{t-p} + \varepsilon_t \quad (23)$$

Las series presentan tendencia en niveles y también tendencia en el largo plazo.

5. Intercepto y tendencia en el vector de cointegración y en el VAR

$$\begin{aligned} \Delta Z_t &= \mu_1 + \mu_{2t} + \alpha\beta' \binom{Z_{t-1}}{t-1} + \Gamma_1 \Delta Z_{t-1} + \\ &\quad \ldots + \Gamma_p \Delta Z_{t-p} + \Gamma t + \varepsilon_t \end{aligned} \quad (24)$$

Las series presentan tendencia en niveles, pero de forma cuadrática. También presenta tendencia en la relación del largo plazo.

3. RESULTADOS

Para estimar el modelo de cointegración se tomará como ejemplo las trayectorias para el producto interno bruto (PIB) y el nivel de exportaciones tradicionales en Colombia, comprendidos para periodos trimestrales entre de enero de 2000 y el tercer trimestre de 2013.

Debido a la diferencia de escalas se trabajará con las variables en logaritmos.

Se puede observar en la figura ?? que aunque el comportamiento estocástico de las variables es diferente, existe una tendencia estable entre ellas. A continuación con el proceso de cointegración se establecerá si este comportamiento se debe a una relación en el largo plazo o si por el contrario se debe a una dependencia espuria.

3.1 *Prueba de Raíz Unitaria y Órdenes de Integración*

Según el procedimiento expresado en la metodología, se debe establecer en primer lugar que las variables estudiadas presentan el mismo orden de integración, lo cual se lleva a cabo a partir de la prueba de raíz unitaria de Dickey-Fuller aumentada.

Tabla 2. Logaritmo del PIB

Value test-statistic	**-2,51**	1,66	3,77
Critical values:	1 %	5 %	10 %
τ	**-4.04**	**-3.45**	**-3.15**
phi2	6.50	4.88	4.16
phi3	8.73	6.49	5.47

Tabla 3. Logaritmo de las EXPORTACIONES

Value test-statistic	**-3,35**	5,48	6,04
Critical values:	1 %	5 %	10 %
τ	**-4.04**	**-3.45**	**-3.15**
phi2	6.50	4.88	4.16
phi3	8.73	6.49	5.47

Con los resultados de las tablas 2 y 3, se concluye que el PIB, presenta al menos una raíz ya que no se rechaza la hipótesis nula a cualquier nivel de significancia, $\tau_{cal} = -2.51 > \tau_{cal}(1\,\%, 5\,\%, 10\,\%)$

Las EXPORTACIONES presentan evidencia de al menos una raíz unitaria a niveles de significancia del 1 % y 5 %, $\tau_{cal} = -3.35 > \tau_{cal}(1\,\%, 5\,\%)$

Para confirmar que efectivamente las series PIB y EXPORTACIONES tienen el mismo orden de integración, se vuelven a calcular las pruebas de raíz unitaria sobre las series diferenciadas.

De acuerdo con los resultados observados en las tablas 4 y 5, se concluye que el valor calculado del estadístico τ es menor que el valor crítico a cualquier nivel de significancia, por tanto se rechaza la hipótesis nula, con lo cual la serie diferenciada del PIB y la serie diferenciada de las EXPORTACIONES son estacionarias y no presentan evidencia de tener raíz unitaria.

Entonces se concluye que $LogPIB \sim I(1)$, y $LogEXP \sim I(1)$, cumpliéndose de esta manera la primera condición para estimar una posible relación de cointegración entre las variables.

Tabla 4. Diferencia del Logaritmo del PIB

Value test-statistic	**-4,64**	7,20	10,77
Critical values:	1 %	5 %	10 %
τ	**-4.04**	**-3.45**	**-3.15**
phi2	6.50	4.88	4.16
phi3	8.73	6.49	5.47

Tabla 5. Diferencia del Logaritmo de las EXP

Value test-statistic	**-5,14**	8,82	13,22
Critical values:	1 %	5 %	10 %
τ	**-4.04**	**-3.45**	**-3.15**
phi2	6.50	4.88	4.16
phi3	8.73	6.49	5.47

Tabla 6. Criterios de selección

	AIC(n)	HQ(n)	SC(n)	FPE(n)	
	1	1	1	1	
			Rezagos		
	1	2	3	4	5
AIC(n)	-1,39E+07	-1,38E+07	-1,38E+07	-1,37E+07	-1,36E+07
HQ(n)	-1,38E+07	-1,36E+07	-1,35E+07	-1,35E+07	-1,33E+07
SC(n)	-1,36E+07	-1,34E+07	-1,31E+07	-1,30E+07	-1,27E+07
FPE(n)	9,39E-01	1,00E+00	1,06E+00	1,08E+006	1,26E+00

3.2 *Modelo del Vector Autoregresivo (VAR): Relación del corto plazo*

Ya sabiendo que las variables presentan el mismo orden de integración, se procede a establecer el número óptimo de rezago del VAR.

Como se puede observar en la tabla 6 el modelamiento del corto plazo viene determinado por un $VAR(1)$, ya que los cuatro criterios de información indican que sólo se debe utilizar un rezago.

La ecuación del modelo del corto plazo está determinada por:

$$dlpib = dlpib.l - 1 + dlexp.l - 1 + const$$

$$\begin{array}{ccc} dlpib.l-1 & dlexp.l-1 & const \\ 1.032037472 & -0.006447295 & -0.310364028 \end{array}$$

$$dlexp = dlpib.l - 1 + dlexp.l - 1 + const$$

$$\begin{array}{ccc} dlpib.l-1 & dlexp.l-1 & const \\ 1.1425620 & 0.7346291 & -11.1541274 \end{array}$$

Aunque el modelo $VAR(1)$ del corto plazo, es fundamental para establecer la relación estructural de la cointegración, lo que trasciende en este método es hallar e interpretar la relación del largo plazo expresada en el vector de cointegración.

3.3 *Especificación del modelo*

De existir un vector de cointegración, los residuos del modelo deben tener un comportamiento ruido blanco para que la relación entre las variables no sea espuria.

Para tal efecto se debe demostrar que los residuos no se encuentran correlacionados y que presentan varianza homogénea en el tiempo.

Tabla 7. Test Portmanteau (asymptotic)

data: Residuals of VAR object p1ct
$Chi-squared = 60.6904, df = 60, p-value = 0.4508$

Tabla 8. ARCH (multivariate)

data: Residuals of VAR object p1ct
$Chi-squared = 44.6751, df = 45, p-value = 0.4856$

En la tabla 7 se puede observar que la prueba de autocorrelación residual (test de Pormanteau), indica que no existe suficiente evidencia para rechazar la hipótesis de nula, con lo cual se concluye que no existe evidencia de correlación entre los residuales del modelo planteado.

En la tabla 8 se puede observar que la prueba de heteroscedasticidad multivariada (ARCH multivariate), indica que no existe suficiente evidencia para rechazar la hipótesis de nula, con lo cual se concluye que no existe evidencia de heteroscedasticidad temporal entre los residuales del modelo planteado.

3.4 *Modelo de Cointegración (VEC): Relación del largo plazo*

Un modelo de cointegración debe seguir este procedimiento para su estimación:

1. Estimar e interpretar los vectores de cointegración y la relación del corto plazo.
2. Comprobar la distribución de los residuales.
3. Comprobar la existencia de al menos una relación de cointegración a través de la prueba de la traza o del máximo valor propio.

Para esto se debe partir de los siguientes supuestos:

- El vector z_t de n series $I(1)$, sigue una representación $VAR(p)$ y admite una representación de corrección de error VEC dado por 16
- Los errores del modelo presentan un comportamiento ruido blanco de forma $\varepsilon_t \sim RB(0, \Sigma)$.
- El número de relaciones de cointegración es r y que Π se puede escribir como $-\alpha\beta'$.

Se evalúan las dos pruebas vistas en la sección anterior para determinar el rango o número de relaciones de cointegración:

Tabla 9. Test de cointegración de la traza

	test	10pct	5pct	1pct
$r <= 1$	9.77	7.52	9.24	12.97
$r = 0$	32.54	17.85	19.96	24.60

Tabla 10. Test de cointegración del valor propio

	test	10pct	5pct	1pct
$r <= 1$	9.77	7.52	9.24	12.97
$r = 0$	42.31	13.75	15.67	20.20

- La prueba de la traza: Esta prueba examina la hipótesis nula de que el rango de cointegración es r contra la hipótesis alterna de que el rango es n.
- La prueba del máximo valor propio: Esta prueba examina la hipótesis nula de que el rango de cointegración es r contra la hipótesis alterna de que el rango es $r + 1$.

Las pruebas de la traza y del máximo valor propio en las tablas 6 y 7, muestran resultados similares en lo correspondiente al número de vectores de cointegración existentes en la relación entre el PIB y las EXPORTACIONES.

Para el caso en el cual se prueba la hipótesis nula donde $r = 0$, es decir no existe una relación de cointegración entre el PRODUCTO y las EXPORTACIONES, esta es rechazada a cualquier nivel de significancia, puesto que el valor calculado en la prueba ($traza = 32.54$, $\text{valor propio} = 42.31$) es mayor al valor crítico a cualquier nivel de significancia, es decir, existe un vector de cointegración que explica la relación a largo plazo entre las variables. Este resultado se puede observar en la segunda fila de las tablas 9 y 10.

Ahora cuando se evalúa la hipótesis nula de la existencia de al menos una relación de cointegración en el largo plazo entre las variables, $r \leq 1$, esta no se rechaza a un nivel de significancia del 1 %. Este resultado se encuentra en la primera fila de las tablas 9 y 10.

Para la estimación del vector de cointegración, se elige el componente determinístico de la ecuación 22, empíricamente es la mejor aproximación al comportamiento de las series, ya que asume que la relación en el largo plazo entre las variables fluctúa alrededor de una constante,

Tabla 11. Vector de Cointegración:

	Coint
$LogPIB$	1
$LogEXP$	-0.2201712
C	-9,6425238

$$LogPIB_{t-1} = 0.220171 LogEXP_{t-1} + 9,642524$$

Con lo que se interpreta que un crecimiento de un 1 % en las exportaciones genera un incremento del PIB en un 0.22 %.

4. CONCLUSIONES

- En la aplicación se pudo observar que tanto las series del producto interno bruto (PIB) y las exportaciones tradicionales tienen grado de integración de orden uno, por esta razón es posible tratar de encontrar al menos en teoría un vector cointegración que explique las relaciones de las mismas en el largo plazo.
- Se encuentra que la relación del corto plazo entre el producto interno bruto (PIB) y las exportaciones tradicionales (EXP) está determinado estructuralmente por un vector autoregresivo de orden 1, $VAR(1)$, lo cual indica que la relación entre las variables en diferencias sólo se remite al primer rezago.
- Para demostrar que la relación de equilibrio del largo plazo que se establece entre las variables a partir del vector de cointegración no es espuria, se demuestra que la distribución de los residuos del modelo presenta un comportamiento ruido blanco a partir del test para evaluar autocorrelación de Portmanteu y de la prueba ARCH multivariada para evaluar la homogeneidad temporal de la varianza en los residuales. En ambos casos no se rechaza la hipótesis nula con lo que se concluye que no existe evidencia de autocorrelación ni de heterocedasticidad en los residuos.
- Se asume que la relación en el largo plazo entre las variables fluctúa alrededor de una constante, y se demuestra a partir de la prueba de la traza y del máximo valor propio que estas series se encuentran cointegradas y que existe un vector de cointegración que dice que un crecimiento de un 1 % en las exportaciones genera un incremento del PIB en un 0.22 %.
- Este es un resultado importante en política económica, ya que incentivaría decisiones gubernamentales que conlleven al aumento en las exportaciones para generar de esa forma un crecimiento sostenido en el producto interno bruto (PIB), el cuál es la principal referencia de crecimiento económico usado por los países.

REFERENCIAS

[Alonso and Patiño, 2007] Alonso, J. C. and Patiño, C. I. (2007). ¿ crecer para exportar o exportar para crecer?, el caso del valle del cauca. *Banco de la república, ensayos sobre economía regional*, (46).

[Brooks, 2002] Brooks, C. (2002). *Introductory Econometrics for Finance*. Cambridge University Press, 6 edition.

[Enders, 2004] Enders, W. (2004). *Applied econometric time series*. John Wiley & Sons.

[Engle and Granger, 1987] Engle, R. F. and Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, pages 251–276.

[Mauricio, 2007] Mauricio, J. A. (2007). *Introducción al análisis de series temporales*. Universidad complutense de Madrid, 1 edition.

[Pfaff, 2006] Pfaff, B. (2006). *Analysis of integrated and cointegrated time series with R*. Springer Science & Business Media.

[Pfaff, 2008] Pfaff, B. (2008). Var, svar and svec models: Implementation within r package vars. *Journal of Statistical Software*, 27(4):1–32.

[Pfaff and Stigler, 2013] Pfaff, B. and Stigler, M. (2013). urca: Unit root and cointegration tests for time series data. *R package version*, pages 1–2.

[Rendón, 2007] Rendón, L. (2007). Causalidad temporal entre el producto y las exportaciones para colombia: Análisis sectorial. *Revista sociedad y economía*, (12).