



BANCO CENTRAL DE RESERVA DEL PERÚ

Análisis de la Versión Débil de la Hipótesis del Mercado Eficiente en el Perú

Freddy Espino*

* Banco Central de Reserva del Perú

DT. N°. 2022-003
Serie de Documentos de Trabajo
Working Paper series
Mayo 2022

Los puntos de vista expresados en este documento de trabajo corresponden a los de los autores y no reflejan necesariamente la posición del Banco Central de Reserva del Perú.

The views expressed in this paper are those of the authors and do not reflect necessarily the position of the Central Reserve Bank of Peru

Análisis de la Versión Débil de la Hipótesis del Mercado Eficiente en el Perú

Freddy Espino¹
Banco Central de Reserva del Perú

Resumen

El presente trabajo analiza la Versión Débil de la Hipótesis de Mercado Eficiente (HME) para el Perú durante el periodo 2006 – 2021, probando empíricamente si el Índice General de la Bolsa de Valores de Lima (IGBVL) muestra una trayectoria similar a la de un proceso estocástico denominado paseo aleatorio. Las pruebas estadísticas indican que dicha característica no se refleja para datos con frecuencia diaria, semanal y mensual, pero sí para datos trimestrales. De esta manera, se concluye que la Versión Débil de la HME no se cumple en el caso peruano.

¹ Jefe del Departamento de Liquidación y Control de Operaciones Internas del Banco Central de Reserva del Perú. El autor agradece los comentarios de Marco Vega. Las opiniones vertidas en este documento son exclusivas del autor y no representan, necesariamente, a las de las instituciones que pertenece.

1. Introducción

El presente trabajo analiza la Versión Débil de la Hipótesis de Mercado Eficiente (HME) para el Perú durante el periodo 2006 - 2021. La prueba empírica consiste en investigar si el Índice General de la Bolsa de Valores de Lima (IGBVL) tiene una trayectoria similar a la de un proceso de paseo aleatorio. Dicho patrón no se reproduce en el caso de datos con frecuencia diaria, semanal y mensual, y sí lo hace para el caso de datos trimestrales. De esta manera, se concluye que la Versión Débil de la HME no se cumple en el caso peruano.

La HME se basa en el trabajo de Fama (1970), quien indica que un mercado es eficiente si los precios de los activos financieros "reflejan plenamente" toda la información disponible. Malkiel (2003) añade que un mercado es eficiente en tanto no permite a los inversionistas obtener rendimientos superiores a la media sin que asuman riesgos superiores que el promedio de inversionistas. En su Versión Débil, la HME señala que los precios de los activos financieros reflejan toda la información pasada del precio, así como toda la información relacionada a su transacción a través del tiempo, como el volumen de transacción, tasas de interés, etc.

En términos estadísticos, la Versión Débil de la HME desemboca en el concepto del proceso estocástico denominado paseo aleatorio, el cual, por definición, no puede ser predecible. La relación entre el concepto de eficiencia de mercado y paseo aleatorio está en que, si los precios de los activos financieros incorporan toda la información disponible, cualquier inversionista no podrá tomar ventaja de la información que maneja, debido a que los demás ya la tienen incorporada en las transacciones que realizan en el mercado, haciendo imposible predecir un precio de tal manera que tome ventaja sobre el resto de los inversionistas. En todo caso, cualquier ganancia extraordinaria será sólo temporal.

Siguiendo las definiciones de paseo aleatorio de Campbell *et. al* (1997), para el caso modelo de paseo aleatorio tipo 1 (RW1)², se realizan las pruebas de (i) secuencias y reversas, (ii) rachas y (iii) de independencia a través de la prueba BDS³; mientras que para el caso del paseo aleatorio tipo 3 (RW3)⁴ se realizan las pruebas de (i) autocorrelación y (ii) ratios de varianza. Los resultados indican que el mercado financiero en el Perú no es eficiente de acuerdo con la Versión Débil de la HME.

El trabajo está organizado en cinco capítulos. En el primero se presenta la introducción, el segundo capítulo expone la teoría de mercados eficientes y una breve revisión de la literatura empírica, el tercero se ocupa de hacer una revisión de la metodología empírica que se emplea en la presente investigación, el cuarto presenta los resultados y el quinto las conclusiones.

² Considera retornos independientes e idénticamente distribuidos. Se desarrolla en el Capítulo 3.

³ En referencia a sus autores Brock, Dechert, Scheinkman y LeBaron (1996).

⁴ Considera retornos no correlacionados. Se desarrolla en el Capítulo 3

2. Marco Teórico

La Hipótesis de Mercado Eficiente (HME) se basa en el trabajo de Fama (1970), quien define un mercado eficiente como aquel en el que los precios proporcionan señales precisas para la asignación de recursos, es decir, uno en el que las empresas puedan tomar decisiones de producción e inversión bajo el supuesto de que los precios en cualquier momento "reflejan plenamente" toda la información disponible.

Por otro lado, Malkiel (2003) añade que un mercado es eficiente en tanto no permite a los inversionistas obtener rendimientos superiores al promedio sin que asuman riesgos superiores a la media, es decir, no hay manera de que un inversionista obtenga mayor rentabilidad que la del mercado asumiendo el mismo riesgo que el resto de los inversionistas.

La HME tiene tres versiones (Fama; 1970):

- i. **Débil.**- En esta versión los precios de los activos financieros reflejan toda la información pasada del precio, así como toda la información relacionada a su transacción a través del tiempo, como el volumen de transacción, tasas de interés, etc.
- ii. **Semi Fuerte.**- En esta versión los precios de los activos financieros reflejan, además de lo indicado en la Versión Débil, todos los fundamentos del precio, como son los factores de demanda y de oferta, así como las variables relacionadas a la empresa, como sus balances y estados de ganancias y pérdidas.
- iii. **Fuerte.**- En esta versión, además de incorporar la información indicada en la versión semi fuerte, los precios de los activos incorporan la información confidencial, es decir, aquella que sólo un pequeño grupo de personas relacionadas a la empresa pueden obtener.

La HME tiene los siguientes supuestos: (i) no hay costos de transacción en la negociación de activos financieros, (ii) toda la información está disponible sin costo para todos los participantes del mercado, y (iii) todos están de acuerdo en las implicancias de la información tanto para el precio actual y como para las distribuciones de los precios de cada activo financiero en el futuro. De esta forma, el precio actual de un activo financiero "refleja completamente" toda la información disponible. No obstante, si bien estas condiciones son suficientes para la eficiencia del mercado, no son necesarias.⁵

Esta acepción de eficiencia de mercado toma forma con el concepto de Juego Limpio, el cual es uno en el que ninguno de los participantes tiene ventaja sobre los demás, por lo que cualquier resultado derivado del mismo se considera justo (Campbell *et.*

⁵ Por ejemplo, tal como señala Fama (1970), incluso grandes costos de transacción que inhiben el flujo de transacciones no implican en sí mismos que los precios no "reflejen completamente" la información disponible. De manera similar, el mercado puede ser eficiente si "un número suficiente" de inversionistas tienen fácil acceso a la información disponible. Y el desacuerdo entre los inversionistas acerca de las implicaciones de la información dada no implica en sí mismo una ineficiencia del mercado a menos que haya inversionistas que puedan hacer consistentemente mejores evaluaciones de la información disponible que las implícitas en los precios del mercado.

al, 1997). Aplicado al mercado financiero, significa que ningún agente tiene más o mejor información que otro, por lo que, como resultado de las transacciones que se dan en el mercado, los precios terminan reflejando toda la información disponible. En términos matemáticos, esta idea se traslada a la definición de martingala, que es un proceso estocástico que satisface la siguiente condición:

$$E[P_{t+1}|P_t, P_{t-1}, \dots] = P_t$$

Donde P_t representa el precio de un activo en el periodo t . La interpretación es como sigue: el valor esperado condicional del precio para el siguiente periodo P_{t+1} , tomando en cuenta el precio de hoy y toda la información pasada, es igual al precio de hoy P_t . Otra forma de interpretar es que la mejor predicción del precio de mañana, con toda la información histórica, es el precio de hoy. Dicha condición se puede reescribir de la siguiente manera:

$$E[P_{t+1} - P_t|P_t, P_{t-1}, \dots] = 0$$

Así, el valor esperado hoy del retorno del siguiente periodo, expresada en la diferencia entre el precio esperado para el siguiente periodo y el de hoy, condicionado a toda la información pasada, es cero. Dado que el precio es un proceso estocástico el cual tiene la misma probabilidad de subir que de bajar el siguiente periodo, entonces el retorno esperado condicional es cero.

Sin embargo, tal como lo señalan Campbell *et. al* (1997), no es una condición necesaria ni suficiente que se cumpla la condición de martingala para la determinación de los precios de los activos, dado que la evidencia muestra que hay una relación directa entre retorno esperado y riesgo, la cual es la base para la teoría de inversión formalizada desde el trabajo de Markowitz (1952). El punto es que la correlación del retorno actual con los pasados es cero o tan bajo que no puede establecerse un patrón identificable, y por lo tanto, predecible.

Tomando en cuenta el siguiente modelo de paseo aleatorio con intercepto:

$$P_t = \mu + P_{t-1} + \epsilon_t, \quad \epsilon_t \sim IID(0, \sigma^2)$$

Donde ϵ_t es un término de error idéntica e independientemente distribuido (IID), entonces el valor esperado condicional para $t + 1$ es P_t y el retorno, definido como la diferencia de precios:

$$R_t = \Delta P_t = \mu + \epsilon_t$$

Entonces, los retornos del precio de los activos son independientes entre periodos y además tienen una distribución similar a lo largo del tiempo. Así, una manera de probar la HME de manera empírica es a través de la medición del grado de autocorrelación del retorno de los activos y definir si es significativo o no, no sólo desde el punto de vista estadístico, si no del valor de los coeficientes estimados de tal manera que se pueda identificar un patrón predecible.

De esta manera, la relación entre el concepto de eficiencia de mercado y paseo aleatorio está en que si los precios de los activos financieros incorporan toda la información disponible, entonces estos se vuelven impredecibles, debido a que

cualquier inversionista no podrá tomar ventaja de la información que maneja, puesto que los demás ya la habrán incorporado a su conjunto de información, haciendo imposible predecir un precio que le permita obtener ganancias extraordinarias de manera sistemática.

La consecuencia de esto es, tal como lo afirma Bodie *et al.* (2018, p. 334), es que un mercado sería ineficiente si los precios muestran un patrón predecible, lo cual es una afirmación extrema, puesto que existe de alguna manera una relación entre los precios de hoy y los del pasado, en tanto los inversionistas toman en cuenta la información pasada para sus decisiones de inversión. De hecho, los agentes tienden a comportarse de manera similar en momentos de mucho optimismo o pesimismo, periodos en los que se observa tendencias al alza y a la baja respectivamente.

Por otro lado, contrario a lo que postula la HME, en la práctica se observa que algunos inversionistas acumulan ganancias exorbitantes, lo cual depende del tamaño del fondo que administre y de la tecnología que usen para explotar al máximo cualquier diferencia mínima entre el precio de compra y de venta de activos financieros. Al respecto, Lewis (2014) narra eventos sucedidos en el mercado financiero de los Estados Unidos de América luego de la crisis financiera de 2008, en donde el método de transacciones de alta frecuencia fue usado para anticipar las órdenes colocadas por los inversionistas, lo cual se tradujo en ingente cantidad de ganancias para los bancos de inversión involucrados.

Desde el punto de vista teórico, las principales críticas a la HME provienen de Grossman y Stiglitz (1980), quienes mostraron que es imposible que un mercado sea perfectamente eficiente desde el punto de vista de la información, dado que esta es costosa, por lo que los precios no pueden reflejar perfectamente la información que está disponible, puesto que, si lo hiciera, los inversionistas que gastaron recursos en obtenerla y analizarla no recibirían compensación.

A pesar de que el marco teórico no fue desarrollado hasta mediados de la década de 1960 por Samuelson (1965) y Mandelbrot (1966), los trabajos empíricos que prueban que el precio de los activos tiene una evolución similar a la de un paseo aleatorio empezaron desde 1900 con el trabajo de Bachelier (1900), de acuerdo con Fama (1970). Luego vienen un conjunto de trabajos, siendo los más relevantes el de Kendall (1953), a pesar de que no fue su objetivo, pero si encuentra dicho patrón, y el de Fama (1965). Décadas más tarde, Fama (1991) hace una revisión de la literatura empírica sobre la HME en sus tres versiones, y encuentra tanto evidencia a favor como en contra de la HME en sus tres versiones.

En el caso del Perú son pocos los trabajos realizados en ese tema. Delgado y Humala (1997) prueban la Versión Débil de la HME en dos partes: primero, a través del análisis de la estacionariedad, autocorrelación, rendimientos por el día de la semana para explorar posibles características estacionales y volatilidad; segundo, mediante el análisis de cointegración del mercado peruano con otros similares. Los autores hallan evidencia estadística en contra de la HME, pero no concluyen que el mercado peruano sea ineficiente.

Duarte y Mascareñas (2014) prueban la Versión Débil de la HME en los 5 principales mercados bursátiles de Latinoamérica, entre los cuales está Perú. Concluyen que en general, las 5 principales economías latinoamericanas estudiadas han experimentado

una mejora de su eficiencia del mercado en los últimos años, y en el caso de Perú desde 2012. Ruiz y García (2020) también prueban dicha versión para el caso del Mercado Integrado Latinoamericano (MILA), en el cual está el Perú, y encuentran que el mercado peruano no es eficiente, resultado opuesto al de Meneses y Pérez (2020).

3. Metodología

En esta sección se discute la metodología para probar la Versión Débil de la HME. Existen tres versiones sobre el paseo aleatorio, discutidas en Campbell *et. al* (1997), las cuales pasaremos a revisar, así como las pruebas empíricas que se realizan en cada caso.

3.1 Paseo Aleatorio 1 (RW1): Incrementos Independientes e Idénticamente Distribuidos (IID)

La versión simple del paseo aleatorio con errores IID es:

$$P_t = \mu + P_{t-1} + \epsilon_t, \quad \epsilon_t \sim IID(0, \sigma^2)$$

Donde P_t es el precio del activo, μ es el intercepto y ϵ_t representa es el término de error IID con media 0 y varianza constante σ^2 . De esta manera, los retornos del precio del activo se expresan de la siguiente manera:

$$R_t = \mu + \epsilon_t$$

La independencia de los retornos implica que el paseo aleatorio es un proceso de juego limpio, pero más estricto que la martingala: la independencia no implica solo que los incrementos no estén autocorrelacionados, sino que también cualquier función no lineal no lo esté también.

La solución por iteración de la versión simple del paseo aleatorio es:

$$P_t = P_0 + \mu t + \sum_{i=1}^t \epsilon_i$$

Donde P_0 es el precio inicial. De esta manera, si calculamos la media y varianza incondicional obtenemos:

$$E[P_t] = P_0 + \mu t$$

$$Var[P_t] = \sigma^2 t$$

Por lo que el paseo aleatorio es un proceso no estacionario, dado que la media y la varianza se incrementan a través del tiempo t , condición que se mantiene para las otras versiones de paseo aleatorio que se discuten más adelante. El supuesto más simple sobre la distribución de ϵ_t es que sea Normal, por lo que sería un proceso $IIDN(0, \sigma^2)$ equivalente a un movimiento aritmético Browniano, distribuido en una muestra separada por unidades. Si bien este supuesto simplifica varios cálculos, padece del mismo problema que afecta a los retornos normalmente distribuidos: si

la distribución del proceso P_t es Normal, entonces habrá siempre la posibilidad de que $P_t < 0$, lo cual no se ajusta a la realidad. Para evitar dicha posibilidad, se calcula el logaritmo natural de P_t tal que $p_t = \log(P_t)$, por lo que el precio queda definido como:

$$p_t = \omega + p_{t-1} + \epsilon_t, \quad \epsilon_t \sim IIDN(0, \sigma^2)$$

De esta manera, el retorno continuo compuesto de los activos se representa como:

$$r_t = \Delta p_t = \omega + \epsilon_t, \quad \epsilon_t \sim IIDN(0, \sigma^2)$$

La independencia en el incremento implica tanto lineal como no lineal, es decir, que la no autocorrelación se da manera lineal, $Cov[\epsilon_t, \epsilon_{t-k}] = 0$ para todo $k \neq 0$, cómo no lineal, $Cov[\epsilon_t^2, \epsilon_{t-k}^2] = 0$ para todo $k \neq 0$. Esta definición implica que la distribución de los datos se mantiene a lo largo de la muestra, no importa que tan grande sea esta.

Las pruebas empíricas correspondientes a este tipo de paseo aleatorio se discuten a continuación:

3.1.1 Secuencias y reversas

De acuerdo con Campbell *et. al* (1997), se asume la versión geométrica browniana del RW1 en el cual el logaritmo del precio sigue un proceso aleatorio sin intercepto y con errores IID:

$$p_t = p_{t-1} + \epsilon_t, \quad \epsilon_t \sim IID(0, \sigma^2)$$

Se define la función I_t para la siguiente variable aleatoria:

$$I_t = \begin{cases} 1 & \text{si } r_t \equiv p_t - p_{t-1} > 0 \\ 0 & \text{si } r_t \equiv p_t - p_{t-1} \leq 0 \end{cases}$$

La prueba consiste en estimar el número de secuencias, definido como el número de veces consecutivos de retornos con el mismo signo, y reversas, definido como el número de veces consecutivos de retornos con signos opuestos. Específicamente, sea la muestra de $n + 1$ retornos r_1, \dots, r_{n+1} el número de secuencias N_s y el número de reversas N_r debe ser expresada como funciones simples de I_t :

$$N_s \equiv \sum_{t=1}^n Y_t, \quad Y_t = I_t I_{t+1} + (1 - I_t)(1 - I_{t+1})$$

$$N_r \equiv n - N_s$$

Si el logaritmo del precio del activo sigue un proceso de paseo aleatorio sin intercepto con errores IID, y si se añade la restricción de que la distribución de los incrementos de ϵ_t son simétricos, entonces si r_t es positivo no negativo, debe tener la misma probabilidad que un experimento de lanzamiento de moneda al aire, es decir, 0.5 en cada caso. Esto implica que, para cada par de retornos consecutivos, una secuencia y una reversa son igualmente probables, lo cual se refleja en el denominado ratio Cowles-Jones:

$$\widehat{CJ} = \frac{N_s}{N_r} = \frac{N_s/n}{N_r/n} = \frac{\hat{\pi}_s}{1 - \hat{\pi}_s}$$

Donde $\hat{\pi}_s$ es la probabilidad de una secuencia. De esta manera, la ratio debe ser aproximadamente igual a uno. El hecho de que esta ratio exceda el valor de 1 para varios retornos históricos lleva a pensar que existe un patrón identificable (Campbell *et. al*, 1997)

No obstante, asumir que el intercepto es cero puede llevar a una conclusión errónea, dado que mientras el intercepto sea distinto de cero, siempre será el caso de que las secuencias serán más probables que las reversas, porque el intercepto induce a una tendencia en el proceso de los datos. Para observar esto, asumamos ahora que:

$$p_t = \omega + p_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2)$$

Entonces, el indicador I_t no es más un experimento de lanzamiento de moneda al aire dado que está sesgado en la dirección del signo del intercepto:

$$I_t = \begin{cases} 1 & \text{con probabilidad } \pi \\ 0 & \text{con probabilidad } 1 - \pi \end{cases}$$

Donde:

$$\pi \equiv \Pr(r_t > 0) = \Phi\left(\frac{\omega}{\sigma}\right)$$

Donde $\Phi(\cdot)$ representa la función densidad de la distribución Normal. Si el intercepto ω es positivo, entonces $\pi > 0.5$ y si es negativo $\pi < 0.5$. Una especificación más general la ratio CJ sería:

$$CJ = \frac{\pi^2 + (1 - \pi)^2}{2\pi(1 - \pi)}$$

El caso del experimento de lanzamiento de moneda al aire sería cuando $\pi = 0.5$ por lo que la ratio CJ sería igual a uno. Para hacer una comparación formal entre el valor estimado bajo la hipótesis de que es un paseo aleatorio, es decir, que $\widehat{CJ} = 1$, usaremos la distribución asintótica de acuerdo con Campbell *et. al* (1997):

$$\widehat{CJ} \sim N\left(\frac{\pi_s}{1 - \pi_s}, \frac{\pi_s(1 - \pi_s) + 2(\pi^3 + (1 - \pi)^3 - \pi_s^2)}{n(1 - \pi_s)^4}\right)$$

Donde $\pi_s = \pi^2 + (1 - \pi)^2$ Entonces, la prueba estadística consiste en estimar el estadístico \widehat{CJ} bajo la hipótesis nula $H_0: \widehat{CJ} = 1$.

3.1.2 Rachas

La prueba de rachas es una prueba no paramétrica, es decir, no hace supuestos sobre la distribución de los datos bajo estudio, y consiste en determinar si un proceso estocástico tiene una trayectoria tal que no se pueda identificar un patrón de

autocorrelación. Tomando en cuenta el signo del retorno de los activos, bajo la hipótesis de paseo aleatorio, no debería tener un patrón identificable.

De acuerdo con Gujarati y Porter (2009, Cap. 12), se define como racha a la secuencia ininterrumpida de valores positivos o negativos en una serie de tiempo.⁶ Asimismo, se define como la extensión de la racha al número de elementos en ella. De esta manera, examinando cuantas rachas se identifican en una serie de tiempo, se puede derivar una prueba de aleatoriedad de rachas, la cual se elabora de la siguiente manera:

- Sea N el número total de observaciones
- Sea N_1 el número de observaciones positivas
- Sea N_2 el número de observaciones negativas
 - Así $N = N_1 + N_2$
- Sea R el número total de rachas

Entonces, bajo la hipótesis nula (H_0) de que las observaciones del proceso son IID, y asumiendo que $N_1 > 10$ y $N_2 > 10$, el número de rachas se distribuye asintóticamente normal con media y varianza:

$$E(R) = \mu_R = \frac{2N_1N_2}{N} + 1$$

$$Var(R) = \sigma_R^2 = \frac{2N_1N_2(2N_1N_2 - N)}{N^2(N - 1)}$$

De esta manera, se construye el intervalo de confianza al $\alpha\%$ de significancia:

$$Prob[\mu_R - N_{\alpha/2}\sigma_R \leq R \leq \mu_R + N_{\alpha/2}\sigma_R] = 1 - \alpha\%$$

Donde $N_{\alpha/2}$ representa el valor crítico de la distribución normal estándar a un nivel de $\alpha/2\%$ de significancia. De esta manera, si R cae fuera del intervalo, entonces rechazamos la H_0 , mientras que, si cae dentro, fallamos en rechazarla.

3.1.2 Prueba de independencia BDS

Basado en el trabajo de Brock, Dechert, Scheinkman y LeBaron (1996), esta prueba se usa para evaluar si una serie es IID. La independencia puede ser lineal, no lineal o caótica. La idea de la prueba BDS consiste en escoger una distancia entre las observaciones denominada ε . Si las observaciones de la serie son IID, entonces para cualquier par de observaciones la probabilidad de distancia menor o igual a ε debe ser igual desde el punto de vista estadístico.

Sean dos observaciones $\{y_s, y_t\}$ la distancia entre ellas es $|t - s|$ La probabilidad de la distancia entre este par de observaciones se denota como $c_1(\varepsilon)$. A partir de esto, podemos escoger un conjunto de múltiples observaciones que pueden ser consecutivas de acuerdo con la muestra, por ejemplo, dada la observación s y dada

⁶ Sea r_t el retorno de un activo, se define el valor 1 si $r_t > 0$ y cero en otro caso. Entonces, en la secuencia de retornos 0001110001 se identificará dos rachas negativas y dos positivas. En un extremo, sea la secuencia de retornos 01, entonces habrá una racha negativa y otra positiva.

la observación t de una serie y_t , se puede construir el siguiente conjunto de observaciones ordenadas de par en par:

$$\{(y_s, y_t), (y_{s+1}, y_{t+1}), \dots, (y_{s+m-1}, y_{t+m-1})\}$$

Donde m es el número consecutivo de observaciones en del conjunto de dimensión incorporada. Se define la probabilidad conjunta de cada dimensión como $c_m(\varepsilon)$. Entonces, bajo el supuesto de independencia, la probabilidad conjunta del conjunto de dimensión incorporada debe ser el producto de la probabilidad de cada par de observaciones individualmente:

$$c_m(\varepsilon) = c_1^m(\varepsilon)$$

Para una serie y_t con n observaciones la probabilidad de una dimensión particular, referida como la correlación integral, se define como:

$$c_{m,n}(\varepsilon) = \frac{2}{k(k-1)} \sum_{t < s} I_t(y_t^m, y_s^m)$$

Donde $y_t^m = (y_t, y_{t+1}, \dots, y_{t+m-1})$, $y_s^m = (y_s, y_{s+1}, \dots, y_{s+m-1})$, $k = n - m + 1$ e $I_t(y_t^m, y_s^m)$ es una función de tal manera que:

$$I_t(y_t^m, y_s^m) = \begin{cases} 1 & \text{si } |t - s| < \varepsilon \\ 0 & \text{en otro caso} \end{cases}$$

El estadístico BDS se define como:

$$w_{m,n}(\varepsilon) = \frac{[c_{m,n}(\varepsilon) - c_1^m(\varepsilon)]\sqrt{n}}{\sigma_{m,n}(\varepsilon)}$$

Donde $\sigma_{m,n}(\varepsilon)$ es la desviación estándar y puede ser estimado consistentemente. Dicho estadístico se distribuye asintóticamente Normal, bajo la hipótesis de nula de que el proceso es IID.

3.2 Paseo Aleatorio 2 (RW2): Incrementos Independientes

En esta versión se relaja el supuesto de que los incrementos son idénticamente distribuidos a lo largo de la muestra de datos, debido a que es difícil sostener que la distribución de los datos sea la misma a lo largo de una muestra larga, como por ejemplo, 30 o 50 años, sobre todo en economías que sufren constantes cambios en periodos cortos, como las economías en desarrollo.

A pesar de que es menos restrictivo que el RW1, contiene aún la principal característica económica de un paseo aleatorio IID: cualquier transformación arbitraria del incremento del precio en el futuro es impredecible usando cualquier transformación del incremento en el pasado. Las pruebas estadísticas en este caso son:

3.2.1 Reglas de Filtros

De acuerdo con Campbell *et. al* (1997), para probar esta hipótesis se aplica una regla de inversión, que consiste en comprar un activo cuando crece en $x\%$, y vender, incluso en corto, cuando su precio cae $x\%$. La idea es que, si existe tendencias en el mercado de capitales, pero que están ocultos por el rápido movimiento de los precios, entonces, es factible filtrar todos los movimientos pequeños con relación a un nivel y analizar el resto de los movimientos. El retorno total de esta estrategia se considera como prueba de la predictibilidad del retorno de los activos. Se compara el retorno total de este filtro a un portafolio determinado aplicándola al índice compuesto como al IGBVL, y si es mayor, se concluye que existe tendencia en el mercado de capitales.

Sin embargo, los trabajos empíricos que hallan estrategias de inversión que le ganan al mercado, bajo el enfoque de reglas de filtros, lo realizan con datos históricos y asumiendo un costo no coherente con la realidad, por lo que terminan siendo inaplicables en la práctica. De esta manera, consideramos que no es relevante realizar una prueba en ese sentido, dado que cualquier regla que pueda hallarse quedará siempre en el ámbito teórico.

En general, es difícil encontrar una estrategia de inversión que supere al mercado de manera sistemática y que sea publicada, puesto que, al momento de su publicación, ya todos lo habrán incorporado en su conjunto de estrategias, por lo que las ganancias extraordinarias ser irán desvaneciendo. Como indica Bodie *et. al* (2018, p. 347), no es posible encontrar una estrategia como tal publicada dado que el creador estaría perdiendo dinero al hacerlo.⁷

3.2.2 Análisis Técnico

El análisis técnico es esencialmente la búsqueda de patrones recurrentes y predecibles en los precios de los activos (Bodie *et. al*; 2018). Se centra en identificar patrones de movimientos de precios, volúmenes de transacción, señales de negociación y varios otros elementos a través del análisis gráfico para evaluar la fortaleza o debilidad de un activo financiero. Los patrones clásicos que tratan de encontrar son doble profundidad, cabeza y hombros, nivel de apoyo y resistencia, etc.⁸ Aunque dicho enfoque no tiene mucho apoyo desde el punto de vista académico, es muy usado por los inversionistas, por lo que viene ganando atención por parte de los investigadores (Campbell *et al.*, 1997).

No obstante, al igual que la Regla de Filtros, este esquema se basa en un exhaustivo análisis de datos históricos hasta encontrar una estrategia que sea rentable durante la muestra analizada, pero nada garantiza que eso funcione en el futuro o que sea aplicable. En caso funcione al inicio, ocurrirá lo mismo que la Regla de Filtros: los inversionistas lo tomarán en cuenta en su conjunto de información, haciendo que las

⁷ Tal como señalan Bodie *et al.* (2018, p. 347) “Este es un problema de sesgo de selección; los resultados que podemos observar han sido preseleccionados a favor de intentos fallidos. Por lo tanto, no podemos evaluar de manera justa la verdadera capacidad de los administradores de cartera para generar estrategias ganadoras en el mercado de valores.” En el caso de estudio, no se podría probar dicha hipótesis de manera objetiva.

⁸ Una referencia importante es La Teoría de Olas de Elliot (*Elliott Wave Theory*).

ganancias extraordinarias desaparezcan paulatinamente. De esta manera, en este trabajo tampoco se realiza una prueba en ese sentido.

3.3 Paseo Aleatorio 3 (RW3): Incrementos No Correlacionados

De acuerdo con Campbell *et al.* (1997), bajo la Versión Débil del paseo aleatorio, RW3, la primera diferencia de los niveles del paseo aleatorio no está correlacionado para todos los rezagos o adelantos. Esta es la Versión Débil del paseo aleatorio que contiene a las versiones RW1 Y RW2 como casos especiales. Entonces, debemos probar la hipótesis nula de que el coeficiente de autocorrelación es cero para varios rezagos.

Un ejemplo particular sobre un proceso que cumple con el supuesto RW3, pero no con RW1 y RW2, es cualquier proceso para el cual $Cov[\epsilon_t, \epsilon_{t-k}] = 0$ para todo k , pero $Cov[\epsilon_t^2, \epsilon_{t-k}^2] \neq 0$ para algún $k \neq 0$. Este proceso tiene incrementos no correlacionados, pero no son independientes, ya que el cuadrado de los incrementos está correlacionado (efectos ARCH).

3.3.1 Prueba de Autocorrelación

El coeficiente de autocorrelación es una medida de fácil implementación e interpretación. Dada una serie estacionaria, como el retorno de los activos, $\{r_t\}$, la autocovarianza $\gamma(k)$ y la autocorrelación $\rho(k)$ de orden k se definen como:

$$\gamma(k) = Cov[r_t, r_{t-k}]$$

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)}$$

Notar que $\gamma(0)$ es la varianza de la serie $\{r_t\}$. Dado que RW1 implica que todas las autocorrelaciones sean cero, la prueba estadística más simple, bajo la hipótesis nula de que la serie es RW1, que tiene poder bajo varias hipótesis alternativas, es el estadístico Q de Box y Pierce (1970) que se distribuye con m grados de libertad:

$$Q_m \equiv T \sum_{k=1}^m \rho^2(k) \sim \chi_m^2$$

Ljung y Box (1978) proveen el siguiente estadístico corregido para muestras finitas:

$$Q'_m \equiv T(T+2) \sum_{k=1}^m \frac{\rho^2(k)}{T-k} \sim \chi_m^2$$

Sumando todas las autocorrelaciones al cuadrado, la prueba Box-Pierce está diseñada para identificar coeficientes de autocorrelación distintos de cero en cualquier dirección y para cualquier rezago. Por lo tanto, tiene poder contra varias hipótesis alternativas (estadístico *portmanteu*) de ruido blanco. Sin embargo, seleccionar el número de rezagos requiere mucho cuidado: si muy pocos son usados, se puede estar obviando órdenes de autocorrelación mayores, mientras que, si

muchos son usados, la prueba pierde poder. De esta manera, si bien esta prueba tiene bastante aceptación en la práctica, debe haber mejores pruebas estadísticas en donde se fije claramente la hipótesis alternativa.

3.3.2 Ratios de Varianza

Una de las propiedades del modelo RW1 es que la varianza se incrementa a través del tiempo. Por ejemplo, bajo la hipótesis nula de RW1, donde el retorno continuo compuesto $r_t \equiv p_t - p_{t-1}$ es IID, la varianza de $r_t + r_{t+1}$ debe ser dos veces la varianza de r_t . Por lo tanto, probar si una secuencia se asemeja a un paseo aleatorio, debe ser el caso de que la varianza de $r_t + r_{t+1}$ debe ser dos veces la varianza de $\{r_t\}$. Bajo la hipótesis nula de RW1, estos deben ser iguales, por lo que se construye la ratio de varianzas.

Sea $r_t(2) = r_t + r_{t+1}$, entonces se construye el estadístico:

$$VR(2) = \frac{Var[r_t(2)]}{2 \times Var[r_t]}$$

Dicha ratio se puede descomponer de la siguiente manera:

$$VR(2) = \frac{Var[r_t + r_{t+1}]}{2 \times Var[r_t]} = \frac{2 \times Var[r_t] + 2 \times Cov[r_t, r_{t+1}]}{2 \times Var[r_t]} = 1 + \rho(1)$$

Donde $\rho(1)$ es el coeficiente de autocorrelación de primer orden. Para cualquier serie estacionaria, el valor poblacional de la ratio de varianza $VR(2)$ es simplemente uno más el coeficiente de autocorrelación de primer orden. Entonces, bajo la hipótesis nula RW1 el coeficiente de autocorrelación es cero, por lo que $VR(2) = 1$. En la presencia de autocorrelación, $VR(2)$ será mayor a uno para el caso de autocorrelación positiva, por lo que la varianza crecerá rápidamente. De manera alterna, si el coeficiente de autocorrelación es negativo, la varianza va a crecer lentamente.

Para comparaciones mayores a un periodo, la ratio de varianza se representa como:

$$VR(q) = \frac{Var[r_t(q)]}{q \times Var[r_t]} = 1 + 2 \times \sum_{k=1}^{q-1} \left(1 - \frac{k}{q}\right) \times \rho(k)$$

Donde $r_t(k) = r_t + r_{t+1} + \dots + r_{t-k+1}$ y $\rho(k)$ es el coeficiente de autocorrelación de orden k . Esto muestra que el estadístico $VR(q)$ es una combinación lineal particular de los primeros $k - 1$ coeficientes de correlación con pesos linealmente decrecientes. Bajo la hipótesis RW1, la ratio $VR(q)$ debe aproximarse a uno, dado que $\rho(k) = 0$. Dicho resultado se mantiene bajo la hipótesis RW2 y RW3 dado que la varianza de r_t es finita y la varianza promedio converge a un número positivo.

La ratio $VR(q)$ bajo la hipótesis nula de que el retorno es RW1:

$$H_0: r_t = \omega + \varepsilon_t, \quad \varepsilon_t \sim IIDN(0, \sigma^2)$$

No obstante, es un hecho estilizado de que la volatilidad del retorno de los activos no es constante, tal como lo discute Bollerslev *et al.* (1994), por lo que rechazar la hipótesis RW1 bajo la presencia de heteroscedasticidad no es de mucho interés. Entonces, lo que se busca es una prueba estadística bajo la hipótesis nula de que el proceso es un RW3. Mientras los retornos no estén autocorrelacionados, aún en la presencia de heteroscedasticidad, la ratio de varianza se debe aproximar a uno conforme el número de observaciones se incremente ilimitadamente, esto es, la varianza de la suma de los retornos debe ser igual a la suma de la varianza de los retornos. Sin embargo, la varianza asintótica dla ratio de varianza va a depender claramente del tipo de heteroscedasticidad, por lo que la ratio de varianza se construye considerando una forma general de esta.

De esta manera, la prueba desarrollada por Lo y MacKinlay (1988) tiene dos hipótesis: la primera, asume que la serie $\{r_t\}$ es un proceso RW1 con errores IID asumiendo homoscedasticidad; y la segunda, asume que es un proceso RW3 asumiendo heteroscedasticidad. Se define los estimadores de la media de la primera diferencia y la varianza de la diferencia de orden q :

$$\hat{\omega} = \frac{1}{T} \sum_{t=1}^T (p_t - p_{t-1})$$

$$\hat{\sigma}^2(q) = \frac{1}{Tq} \sum_{t=1}^q (p_t - p_{t-q} - q\hat{\omega})^2$$

Y se construye la correspondiente ratio de varianza:

$$VR(q) = \frac{\hat{\sigma}^2(q)}{\hat{\sigma}^2(1)}$$

El estimador de la varianza se corrige por sesgo, tal como lo sugiere Lo y MacKinlay (1988), reemplazando T por $(T - q + 1)$ en el caso que no se considere intercepto, o por $(T - q + 1)(1 - q/T)$ para el caso en el que se le considere. Lo y MacKinlay (1988) muestran que la ratio tiene una distribución asintóticamente normal para un apropiado estimador de $\hat{s}(q)$:

$$z(q) = \frac{VR(q) - 1}{\hat{s}(q)} \sim N(0,1)$$

Bajo la hipótesis de RW1, $\hat{s}(q)$ se define como:

$$\hat{s}^2(q) = \frac{2(2q - 1)(q - 1)}{3qT}$$

Y bajo la hipótesis de RW3 o martingala en diferencia, se usa el estimador de Kernel:

$$\hat{s}^2(q) = \sum_{j=1}^{q-1} \left(\frac{2(q-j)}{q} \right)^2 \delta_j$$

$$\hat{\delta}_j = \frac{\sum_{t=j+1}^T (p_{t-j} - \hat{\mu})^2 (p_t - \hat{\mu})^2}{\left[\sum_{t=j+1}^T (p_{t-j} - \hat{\mu})^2 \right]^2}$$

4. Resultados

En esta sección se presenta los resultados del análisis de datos y de las pruebas realizadas bajo la hipótesis nula de que el retorno del IGBVL es un paseo aleatorio. Para ello, se considera datos diarios desde enero de 2006 hasta setiembre de 2021, extraídos del portal web del Banco Central de Reserva del Perú. El análisis de datos y las pruebas se realizan para datos diarios, semanales, mensuales y trimestrales, con la finalidad de tratar de identificar patrones particulares para cada frecuencia de datos.

4.1 Análisis de Datos

En el Anexo 1 se muestra el gráfico del IGBVL para las distintas frecuencias. Se parte de datos con frecuencia diaria para luego estimar los datos semanales, mensuales y trimestrales tomando el valor del índice de los viernes, los de fin de mes y de cada trimestre respectivamente.

Como se observa, en logaritmos el índice muestra una evolución similar a un paseo aleatorio. Con respecto al retorno continuo, en todos los casos se muestra agrupación de volatilidad, no obstante, conforme la frecuencia de datos disminuye, este patrón va diluyéndose, lo cual está influenciado por el número de observaciones. Asimismo, los histogramas muestran agrupación de los datos alrededor de la media además de presentar observaciones atípicas para el caso de los datos diarios, semanales y mensuales, mientras que en el caso de los trimestrales se observa menos dispersión. El mismo patrón se observa en los gráficos cuantil-cuantil, en donde se observa que la distribución de los datos no sería Normal para el caso de frecuencia diaria, semanal y mensual, mientras que los trimestrales, se aproximan a una distribución Normal.

El Cuadro 1 muestra los principales estadísticos del retorno continuo. Como se observa, el valor del *Skewness* coeficiente de asimetría es negativo en todos los casos, lo cual indica que es más probable observar retornos por debajo de la media que por encima. Por otro lado, el valor de la Kurtosis en los datos diarios, semanales y mensuales, nos indica que tiene una distribución leptocúrtica (Kurtosis > 3), es decir, que tiene valores extremos que no serían reproducidos por una distribución Normal; mientras que para el caso de datos trimestrales, dicho valor está cercano a 3, por lo que estaría más cercano a una distribución Normal, lo cual se corrobora en *p-value* de la prueba Jarque-Bera, al calcularse un valor mayor 5% de significancia, por lo que se falla en rechazar la hipótesis nula de que los datos provienen de una distribución Normal.

Cuadro 1: Principales Estadísticos del Retorno del IGBVL

	Daily	Weekly	Monthly	Quarterly
Mean	0.032666	0.163536	0.621668	1.810990
Median	0.016602	0.207665	0.723352	1.481496
Maximum	12.81556	19.31020	32.54098	34.62436
Minimum	-13.29076	-34.60418	-46.64854	-46.73887
Std. Dev.	1.447970	3.618416	8.379358	16.74467
Skewness	-0.540232	-1.102097	-0.580343	-0.433199
Kurtosis	14.36990	16.45226	9.050397	3.413405
Jarque-Bera	22240.37	6317.937	297.3102	2.380665
Probability	0.000000	0.000000	0.000000	0.304120
Sum	133.6677	133.4450	116.8736	112.2814
Sum Sq. Dev.	8577.259	10670.74	13129.95	17103.43
Observations	4092	816	188	62

4.2 Paseo Aleatorio 1 (RW1)

En el Cuadro 2 se muestra los resultados de las pruebas de secuencias y reversas. Tal como se observa, en todos los casos el valor del estadístico CJ está por encima del valor de uno, y no se halla evidencia a favor de la hipótesis nula de que el IGBVL sea un proceso similar a un paseo aleatorio, al rechazar y la prueba bajo la hipótesis nula de que es igual a uno al 5% de significancia, es decir, de que se comporta como un paseo aleatorio tipo RW1, al obtener *p-value* menor a 0.05 en todos los casos, excepto para los datos trimestrales. Por lo tanto, bajo estos resultados, no hallamos evidencia a favor de la HME, excepto para los datos trimestrales.

Cuadro 2: Prueba de Secuencias y Reversas

	Daily	Weekly	Monthly	Quarterly
N	4092	816	188	62
Ns	2228	442	107	36
Nr	1864	374	81	26
CJ	1.195279	1.181818	1.320988	1.384615
μ	0.032666	0.163536	0.621668	1.810990
σ	1.447970	3.618416	8.379358	16.74467
π	0.508999	0.518024	0.529571	0.543063
π_s	0.500162	0.500650	0.501749	0.503709
CJ sd	0.031296	0.070287	0.147401	0.259695
z	6.239809	2.586785	2.177650	1.481028
p-value	0.000000	0.004844	0.014716	0.069300

En el Cuadro 3 podemos ver los resultados de la prueba de rachas. Tal como se observa, se rechaza la hipótesis nula de que los retornos sean IID a un nivel de significancia $\alpha = 5\%$, por lo que no muestra evidencia a favor de la hipótesis de que el IGBVL sea un proceso similar a un paseo aleatorio RW1, excepto para el caso de los datos trimestrales, en el cual fallamos en rechazar la hipótesis nula. Por lo tanto, bajo estos resultados, no hallamos evidencia a favor de la HME, excepto para el caso de los datos trimestrales.

Cuadro 3: Prueba de Rachas

	Daily	Weekly	Monthly	Quarterly
N	4092	816	188	62
N1	2084	435	97	32
N2	2008	381	91	30
Runs mean	2046.294	407.2132	94.90426	31.96774
Runs sd	31.96943	14.21150	6.830298	3.900469
CI Low 5%	1983.634	379.3587	81.51687	24.32282
CI High 5%	2108.954	435.0678	108.2916	39.61266
Runs	1864	374	81	26
	Reject	Reject	Reject	Fail to Reject

En el Cuadro 4 se muestra los resultados de la prueba BDS. En la prueba se considera la distancia en función de la desviación estándar de un valor de 0.5 y de dimensiones 2, 3, 4 y 5.⁹ Tal como se observa, en todos los casos, rechazamos la hipótesis nula de que el IGBVL se comporta como un paseo aleatorio RW1, incluso para los datos trimestrales. Por lo tanto, bajo estos resultados, no hallamos evidencia a favor de la HME.

⁹ En el Anexo 2 se muestra más resultados para distancias en función de la desviación estándar de 1.0 y 1.5. Los resultados nos llevan a las mismas conclusiones: en todos los casos se rechaza la hipótesis nula al 10% de significancia.

Cuadro 4: Prueba BDS

Distance: $\sigma = 0.5$

Daily	Dimension	BDS Statistic	Std. Error	z-Statistic	Prob.
	2	0.016300	0.001126	14.47203	0.0000
	3	0.019288	0.001003	19.23904	0.0000
	4	0.015706	0.000670	23.42649	0.0000
	5	0.010838	0.000393	27.57699	0.0000
Weekly					
	2	0.014498	0.001775	8.167843	0.0000
	3	0.015253	0.001431	10.65874	0.0000
	4	0.010938	0.000866	12.62787	0.0000
	5	0.007165	0.000459	15.59839	0.0000
Monthly					
	2	0.009910	0.003039	3.260827	0.0011
	3	0.009694	0.002313	4.190840	0.0000
	4	0.006744	0.001321	5.103785	0.0000
	5	0.004265	0.000661	6.449054	0.0000
Quarterly					
	2	0.018220	0.005454	3.340684	0.0008
	3	0.014045	0.003859	3.639478	0.0003
	4	0.008831	0.002051	4.305557	0.0000
	5	0.004568	0.000956	4.780335	0.0000

Los resultados de las pruebas bajo la hipótesis de que el IGBVL es RW1 se rechazan en la mayoría de los casos, siendo la excepción el de los datos trimestrales, en el que la prueba de rachas nos indica que los retornos son IID. Por lo tanto, no encontramos evidencia a favor de la Versión Débil de la HME bajo estos resultados.

4.3 Paseo Aleatorio 3 (RW3)

En el Cuadro 5 se muestra los resultados de las pruebas de autocorrelación (AC) de los retornos continuos del IGBVL. Tal como se observa, en la mayoría de los casos, se rechaza la hipótesis nula de que los retornos no están autocorrelacionados, por lo que el IGBVL no se asemejaría a un proceso paseo aleatorio RW3, siendo la excepción los datos trimestrales, aunque en ese caso se rechaza la hipótesis nula en el primer rezago al 5% de significancia. Por lo tanto, no hallamos evidencia a favor de la Versión Débil de la HME, excepto para los datos trimestrales.

En el Gráfico 1 se muestra la función de autocorrelación de los retornos hasta el rezago 24 de manera continua y las bandas punteadas representan el nivel de significancia de 5% de los coeficientes de correlación construido como $\pm 1.96/\sqrt{T}$ donde T es el número de observaciones, por lo que aquellas barras que sobre salgan de esa banda de confianza se consideran como significativos estadísticamente hablando. Como se observa, se identifica coeficientes de correlación significativos, dado indicio de que hay posibles patrones a identificar.

Cuadro 5: Prueba de Autocorrelación

Daily	Lag	AC	Q-Stat	Prob
	1	0.122	60.485	0.000
	3	0.050	71.369	0.000
	6	-0.022	105.91	0.000
	12	0.039	123.10	0.000
	24	0.007	147.99	0.000
<hr/>				
Weekly				
	1	0.127	13.186	0.000
	3	0.059	26.915	0.000
	6	0.027	27.947	0.000
	12	0.014	62.709	0.000
	24	-0.039	75.183	0.000
<hr/>				
Monthly				
	1	0.167	5.3467	0.021
	3	0.110	20.878	0.000
	6	0.027	23.646	0.001
	12	-0.036	30.905	0.002
	24	-0.054	39.358	0.025
<hr/>				
Quarterly				
	1	0.269	4.6947	0.030
	3	-0.111	5.8152	0.121
	6	-0.103	6.9856	0.322
	12	-0.020	13.190	0.355
	24	-0.028	17.988	0.804

Una manera rápida de identificar posibles patrones es mediante la estimación de modelos ARIMA, los cuales explotan la autocorrelación de las series de tiempo para poder realizar predicciones, bajo el supuesto de que los errores no estén autocorrelacionados. Para ello, escogemos el mejor modelo ARIMA en cada caso, a partir de un modelo ARMA(3, 3) usando como criterio de selección de rezagos al de Schwarz:¹⁰

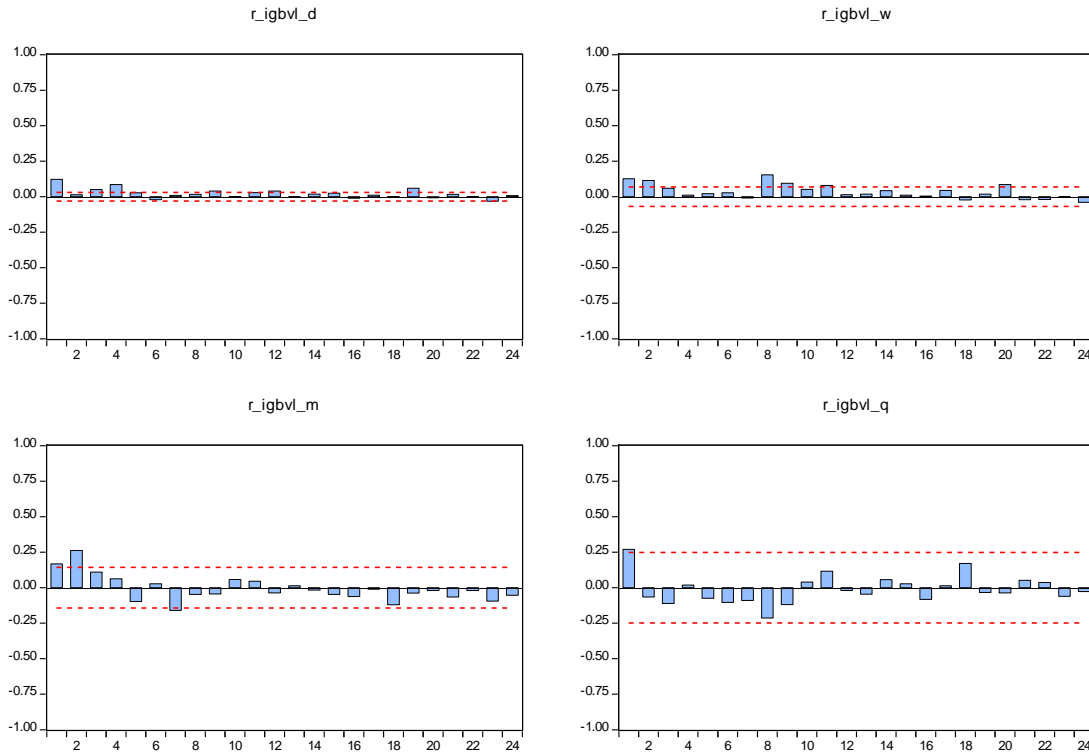
- i. En datos diarios, el modelo fue un ARMA(1, 2), pero se rechaza la hipótesis nula de que los errores no estén autocorrelacionados.
- ii. En datos semanales, el modelo fue un ARMA(1, 1) y se encontró que los errores no estarían autocorrelacionados.
- iii. En datos mensuales, el modelo fue un ARMA(2, 0) y se encontró que los errores no estarían autocorrelacionados.
- iv. En datos trimestrales no se probó, debido a que la función de autocorrelación de los retornos no están autocorrelacionados.

¹⁰ Los detalles están en el Anexo 3.

Como se observa en dos de los tres casos posibles se haya un modelo ARMA que reúne las condiciones necesarias para realizar predicción, y una búsqueda exhaustiva de un mejor modelo ARIMA debería arrojar mejores resultados en hallar un patrón.

Gráfico 1

Autocorrelation Function



En el Cuadro 6 se observa los resultados de la razón de varianza de Lo y MacKinlay (1988), asumiendo incrementos heteroscedásticos en el paseo aleatorio, una media distinta de cero y corrigiendo por el sesgo de varianza. Como se observa, en la mayoría de los casos rechazamos la hipótesis nula de que el IGBVL se comporta como un RW3 al 5% de significancia, aunque, como en los casos anteriores, para el caso de datos trimestrales hay evidencia a favor de que el índice sea RW3, y en menor medida en el caso de datos mensuales cuando se compara la varianza de dos periodos consecutivos. Por lo tanto, no hallamos evidencia a favor de la Versión Débil de la HME, excepto para los datos trimestrales.

Como se observa, en todas las pruebas no se encuentra evidencia a favor de la HME, excepto para los datos trimestrales. De esta manera, a la luz de los resultados, se puede concluir que el mercado financiero en el Perú no es eficiente de acuerdo con la Versión Débil de la HME.

Cuadro 6: Prueba de Ratio de Varianza

Daily	Period	Var. Ratio	Std. Error	z-Statistic	Probability
	2	1.120753	0.039575	3.051230	0.0023
	3	1.170919	0.058437	2.924827	0.0034
	4	1.221802	0.072697	3.051063	0.0023
	5	1.286584	0.084314	3.399021	0.0007
	6	1.339812	0.094159	3.608905	0.0003
<hr/>					
Weekly					
	2	1.135797	0.057784	2.350063	0.0188
	3	1.261167	0.095905	2.723178	0.0065
	4	1.345533	0.124928	2.765851	0.0057
	5	1.397875	0.146945	2.707639	0.0068
	6	1.443860	0.164583	2.696883	0.0070
<hr/>					
Monthly					
	2	1.167296	0.088855	1.882801	0.0597
	3	1.419031	0.134576	3.113709	0.0018
	4	1.601242	0.169616	3.544714	0.0004
	5	1.738305	0.196761	3.752295	0.0002
	6	1.802845	0.221494	3.624680	0.0003
<hr/>					
Quarterly					
	2	1.279636	0.170683	1.638338	0.1014
	3	1.321888	0.248082	1.297506	0.1945
	4	1.270763	0.303454	0.892272	0.3722
	5	1.203803	0.346371	0.588394	0.5563
	6	1.075430	0.381280	0.197835	0.8432

Dado los resultados, se puede inferir que, si los inversionistas tratan de hallar patrones en los datos diarios, semanales y mensuales, deberían encontrarlos.¹¹ No obstante, esto no implica que al explotarlos vayan a obtener ganancias extraordinarias, puesto que de hecho todos los que actualmente están en el mercado lo vienen haciendo. Por lo tanto, no hay que perder de vista que los resultados hallados sólo se centran en las características de los retornos del mercado bursátil, no en la forma que estos pueden ser explotados, por lo que la prueba empírica de la Versión Débil de la HME queda de alguna manera incompleta.

El hecho de que en datos trimestrales fallemos en rechazar la HME, se puede explicar, en parte, por la forma en que estimamos los datos trimestrales al tomar los datos de fin de cada trimestre, haciendo que los patrones que se puedan hallar en datos de mayor frecuencia se diluyan.

Los resultados, interpretados bajo el enfoque teórico señalado, indican que el mercado de capitales peruanos no es eficiente en el sentido de que los precios de los

¹¹ En realidad, lo más probable es que ya lo vengamos realizando, por lo que la idea va dirigida, principalmente, a los nuevos inversionistas.

activos, a nivel agregado, no estarían reflejando toda la información disponible. En consecuencia, las implicancias de política económica apuntarían a que se deba mejorar el flujo de información en el mercado de capitales peruano, como por ejemplo, un mejor acceso gratuito a los precios históricos de los activos, así como de los balances de las empresas que cotizan en bolsa y toda la información relacionada a la toma de decisiones por parte de los inversionistas, mediante un rediseño de las páginas web de las instituciones que representan el mercado de capitales peruano como son la Bolsa de Valores de Lima y la Superintendencia de Mercado de Valores. Asimismo, se debe crear los incentivos para que mayores empresas residentes coticen en la bolsa, impulsando así el incremento del número de inversionistas, promoviendo que la información fluya entre más participantes.

Definitivamente las medidas propuestas en este trabajo son insuficientes para la mejora en el flujo de información en el mercado de capitales peruano, teniendo en cuenta que no es el objetivo de este, pero consideramos que es una propuesta básica que se deben tener en cuenta para mejorar el acceso a la información con el objetivo de que los precios de los activos reflejen toda la información disponible de acuerdo con los resultados obtenidos.

5. Conclusiones

En el presente trabajo se analizó la Versión Débil de la Hipótesis de Mercado Eficiente (HME) para el Perú durante el periodo 2006 - 2021. La prueba empírica consiste en investigar si el IGBVL tiene una trayectoria similar a la de un proceso de paseo aleatorio. Dicho patrón no se identificó en el caso de datos con frecuencia diaria, semanal y mensual, y sí lo hace para el caso de datos trimestrales. De esta manera, se concluye que la Versión Débil de la HME no se cumple en el caso peruano.

Los resultados, interpretados bajo el enfoque teórico señalado, indican que el mercado de capitales peruanos no es eficiente en el sentido de que los precios de los activos, a nivel agregado, no estarían reflejando toda la información disponible. En consecuencia, las implicancias de política económica apuntarían a que se deba mejorar el flujo de información en el mercado de capitales peruano, como por ejemplo, un mejor acceso gratuito a los precios históricos de los activos, así como de los balances de las empresas que cotizan en bolsa y toda la información relacionada a la toma de decisiones por parte de los inversionistas, mediante un rediseño de las páginas web de las instituciones que representan el mercado de capitales peruano como son la Bolsa de Valores de Lima y la Superintendencia de Mercado de Valores. Asimismo, se debe crear los incentivos para que mayores empresas residentes coticen en la bolsa, impulsando así el incremento del número de inversionistas, promoviendo que la información fluya entre más participantes.

Definitivamente las medidas propuestas en este trabajo son insuficientes para la mejora en el flujo de información en el mercado de capitales peruano, teniendo en cuenta que no es el objetivo de este, pero consideramos que es una propuesta básica que se deben tener en cuenta para mejorar el acceso a la información con el objetivo de que los precios de los activos reflejen toda la información disponible de acuerdo con los resultados obtenidos.

Las siguientes investigaciones deberían estar enfocadas en realizar un análisis similar a nivel desagregado, esto es, por el precio de cada activo financiero que esté registrado en el mercado de capitales peruano. Asimismo, se debería estudiar otras anomalías en el mercado, como el efecto calendario, la relación entre volatilidad y correlación entre los activos, etc. Asimismo, se debe explorar posibles no linealidades en el patrón de los retornos de los activos en dicho mercado.

Bibliografía

Bachelier, Louis (1900) "Théorie de la Speculation", *Annales Scientifiques de l'E.N.S.*, Vol. 17, N°3, pp. 21-86.

Bodie, Z., A. Kane y A. Marcus (2018) *Investments*, New York: Mc Graw Hill Education, 966 pp, 11th Edition.

Bollerslev, T., Engle, R., y D. Nelson (1994) "ARCH Models", *Handbook of Econometrics*, Vol. IV, Ch. 49, pp. 2959 - 3038.

Brock, Dechert, Scheinkman y LeBaron (1996) "A test for independence based on the correlation dimension", *Econometric Review*, Volume 15, N° 3, pp 197-235.

Campbell, J., A. Lo y A. MacKinlay (1997) *The Econometrics of Financial Markets*, Princeton: Princeton University Press, 540 pp.

Delgado, A. y A. Humala (1997) "El mercado bursátil peruano y la hipótesis del mercado eficiente", *Revista de Estudios Económicos*, Banco Central de Reserva el Perú, N° 1, pp. 73 - 92.

Duarte, J. y J. Mascareñas (2014) "Comprobación de la eficiencia débil en los principales mercados financieros latinoamericanos", *Estudios Gerenciales*, Vol. 30, N° 133, pp. 365 - 375.

Fama, Eugene (1965) "The Behavior of Stock-Market Prices", *The Journal of Business*, Vol. 38, No. 1, pp. 34-105

Fama, Eugene (1970) "Efficient Capital Markets: A Review of Theory and Empirical Work", *The Journal of Finance*, Vol. 25, No. 2, pp. 383-417

Fama, Eugene (1991) "Efficient Capital Markets: II", *The Journal of Finance*, Vol. 46, No. 5, pp. 1575-1617

Grossman, S. y J. Stiglitz (1980) "On the Impossibility of Informationally Efficient Markets", *The American Economic Review*, Vol. 70, No. 3, pp. 393-408.

Gujarati, D. y D. Porter (2009) *Basic Econometrics*, New York: McGraw-Hill/Irwin, 946 pp., 5th Edition.

Enders, Walter (2015) *Applied Econometric Time Series*, Alabama: University of Alabama, 498 pp, 4th Edition.

Kendall, Maurice (1953) "The Analysis of Economic Time-Series-Part I: Prices", *Journal of the Royal Statistical Society. Series A (General)*, Vol. 116, No. 1 (1953), pp. 11-34.

Lewis, Michael (2014) *Flash Boys*, New York: W.W. Norton & Company, 320 pp.

Malkiel, Burton (2003) "The Efficient Market Hypothesis and Its Critics", *Journal of Economic Perspectives*, Vol. 17, N° 1, pp. 59 - 82.

Mandelbrot, Benoit (1966) "Forecasts of Future Prices, Unbiased Markets, and "Martingale" Models", *The Journal of Business*, Vol. 39, No. 1, pp. 242-255

Markowitz, Harry (1952) "Portfolio Selection", *The Journal of Finance*, Vol. 7, No. 1, pp. 77-91.

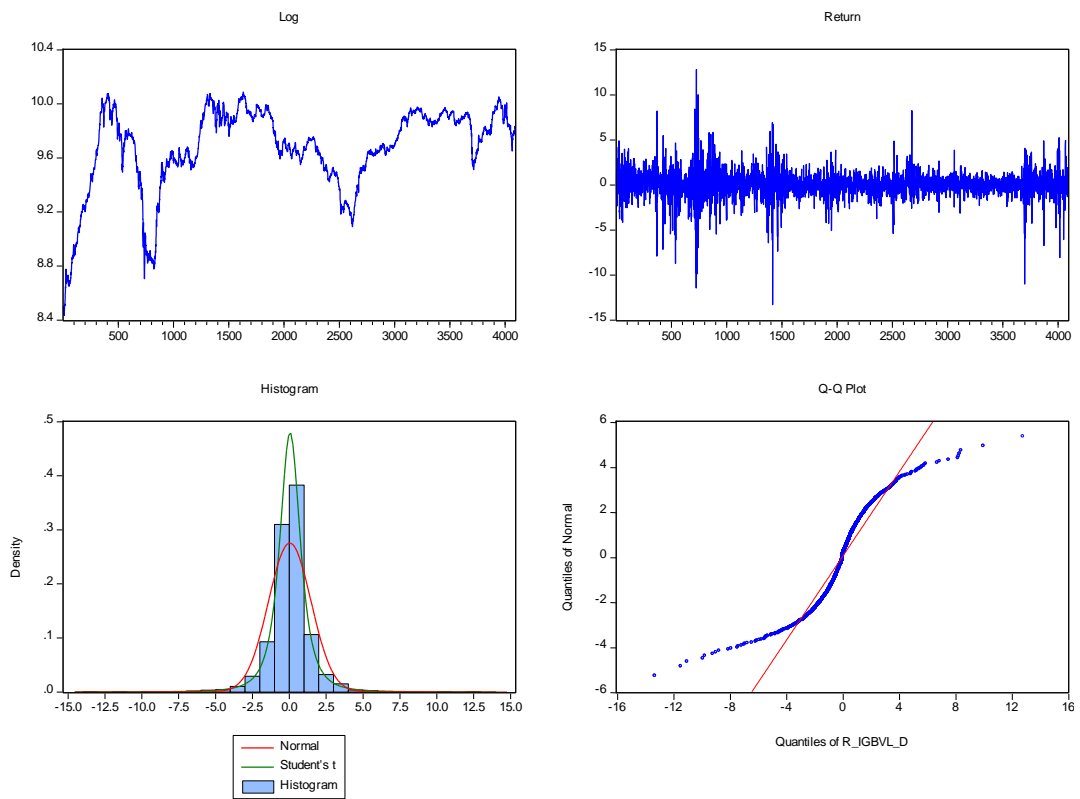
Meneses, L. y C. Pérez (2020) "Análisis comparativo de eficiencia en mercados emergentes. El caso de Colombia, Chile y Perú", *Apuntes Contables*, N° 26, pp. 9 - 24.

Ruiz, B. y G. García (2020) "Hipótesis de Mercados Eficientes y estrategias de inversión en el MILA: 2014-2019", *Análisis Económico*, Vol. XXXV, N° 90, pp. 67-90

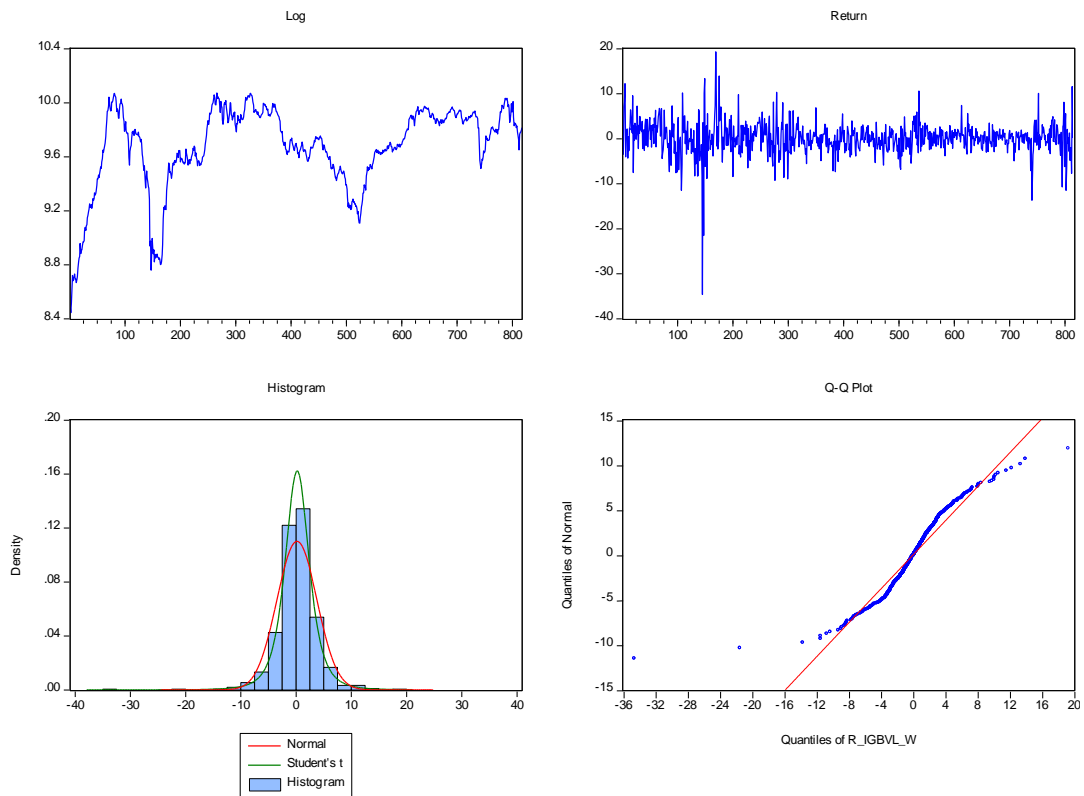
Samuelson, Paul (1965) "Proof That Properly Anticipated Prices Fluctuate Randomly", *Industrial Management Review*, Vol. 6, N° 2, pp. 41 - 49

ANEXO 1

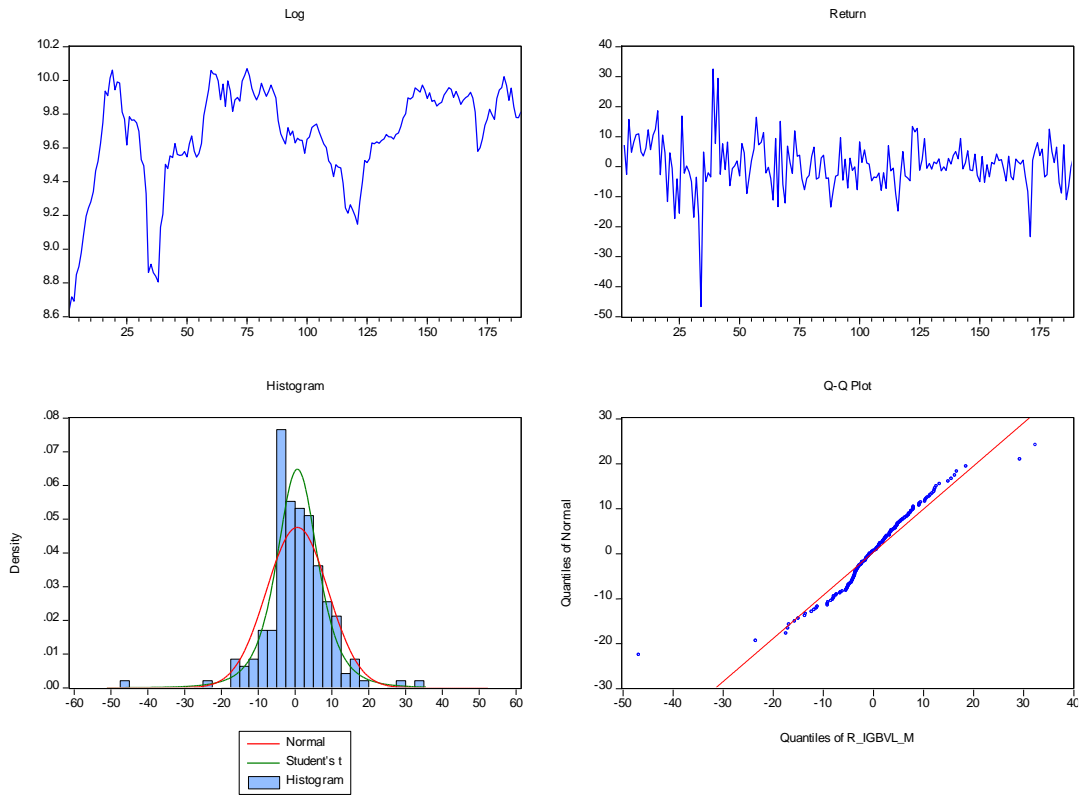
Frecuencia Diaria igbv_d



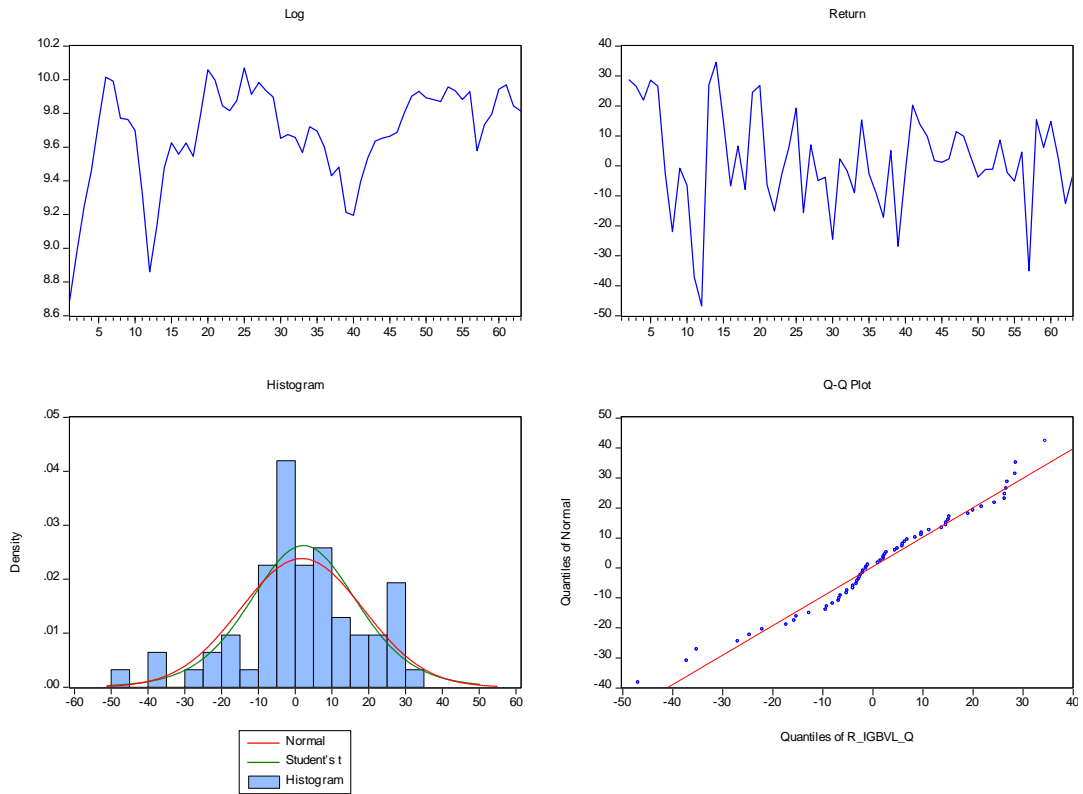
Frecuencia Semanal igbv_w



Frecuencia Mensual igbv_l_m



Frecuencia Trimestral igbv_l_q



ANEXO 2

Prueba BDS

Distance: $\sigma = 1.0$

Daily	Dimension	BDS Statistic	Std. Error	z-Statistic	Prob.
	2	0.031114	0.001671	18.62110	0.0000
	3	0.057907	0.002463	23.51038	0.0000
	4	0.073738	0.002722	27.08699	0.0000
	5	0.079883	0.002634	30.32328	0.0000
Weekly					
	2	0.028839	0.003351	8.607475	0.0000
	3	0.049307	0.004722	10.44101	0.0000
	4	0.059566	0.004991	11.93570	0.0000
	5	0.062946	0.004617	13.63227	0.0000
Monthly					
	2	0.018949	0.006339	2.989477	0.0028
	3	0.033704	0.008615	3.912138	0.0001
	4	0.037587	0.008779	4.281354	0.0000
	5	0.036551	0.007833	4.666456	0.0000
Quarterly					
	2	0.042941	0.009803	4.380474	0.0000
	3	0.048234	0.011956	4.034386	0.0001
	4	0.044990	0.010936	4.113913	0.0000
	5	0.036609	0.008760	4.178922	0.0000

Prueba BDS

Distance: $\sigma = 1.5$

Daily	Dimension	BDS Statistic	Std. Error	z-Statistic	Prob.
	2	0.029381	0.001362	21.57572	0.0000
	3	0.064041	0.002459	26.04458	0.0000
	4	0.094302	0.003326	28.35381	0.0000
	5	0.119087	0.003937	30.24785	0.0000
<hr/>					
Weekly					
	2	0.025107	0.002933	8.559503	0.0000
	3	0.050826	0.005201	9.771856	0.0000
	4	0.074367	0.006909	10.76403	0.0000
	5	0.093111	0.008031	11.59343	0.0000
<hr/>					
Monthly					
	2	0.010205	0.005816	1.754539	0.0793
	3	0.029707	0.010177	2.918948	0.0035
	4	0.044795	0.013339	3.358183	0.0008
	5	0.051789	0.015301	3.384712	0.0007
<hr/>					
Quarterly					
	2	0.043585	0.009738	4.475773	0.0000
	3	0.070380	0.015846	4.441520	0.0000
	4	0.086422	0.019320	4.473280	0.0000
	5	0.086438	0.020618	4.192261	0.0000

ANEXO 3

MODELOS ARMA

Datos Diarios

Dependent Variable: R_IGBVL_D
Method: ARMA Maximum Likelihood (BFGS)
Date: 10/14/21 Time: 15:25
Sample: 2 4093
Included observations: 4092
Convergence achieved after 27 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.034938	0.046256	0.755311	0.4501
AR(1)	0.979770	0.008402	116.6097	0.0000
MA(1)	-0.861205	0.010152	-84.83163	0.0000
MA(2)	-0.099926	0.007713	-12.95517	0.0000
SIGMASQ	2.054378	0.017442	117.7847	0.0000
R-squared	0.019907	Mean dependent var		0.032666
Adjusted R-squared	0.018947	S.D. dependent var		1.447970
S.E. of regression	1.434187	Akaike info criterion		3.560316
Sum squared resid	8406.514	Schwarz criterion		3.568035
Log likelihood	-7279.407	Hannan-Quinn criter.		3.563049
F-statistic	20.75278	Durbin-Watson stat		2.002955
Prob(F-statistic)	0.000000			
Inverted AR Roots	.98			
Inverted MA Roots	.96	-.10		

Datos Diarios Correlograma del Error Estimado

Date: 10/14/21 Time: 17:14

Sample: 1 4093

Included observations: 4092

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.002	-0.002	0.0097	0.922
		2	-0.014	-0.014	0.7596	0.684
		3	0.022	0.022	2.7916	0.425
		4	0.061	0.061	18.014	0.001
		5	0.003	0.004	18.048	0.003
		6	-0.045	-0.044	26.223	0.000
		7	-0.007	-0.010	26.435	0.000
		8	-0.005	-0.010	26.544	0.001
		9	0.021	0.023	28.406	0.001
		10	-0.023	-0.017	30.491	0.001
		11	0.010	0.012	30.862	0.001
		12	0.021	0.019	32.695	0.001
		13	-0.023	-0.025	34.820	0.001
		14	0.002	0.003	34.837	0.002
		15	0.010	0.010	35.272	0.002
		16	-0.029	-0.032	38.742	0.001
		17	-0.001	0.002	38.750	0.002
		18	-0.023	-0.024	40.957	0.002
		19	0.049	0.049	50.747	0.000
		20	-0.029	-0.027	54.267	0.000
		21	0.005	0.007	54.354	0.000
		22	-0.014	-0.015	55.137	0.000
		23	-0.046	-0.053	63.995	0.000
		24	-0.003	-0.003	64.024	0.000
		25	-0.004	0.001	64.100	0.000
		26	-0.017	-0.017	65.270	0.000
		27	0.010	0.019	65.649	0.000
		28	0.042	0.039	72.759	0.000
		29	0.023	0.023	74.986	0.000
		30	-0.035	-0.036	80.033	0.000
		31	-0.027	-0.033	83.013	0.000
		32	0.011	0.007	83.521	0.000
		33	-0.040	-0.045	90.136	0.000
		34	-0.037	-0.029	95.776	0.000
		35	0.010	0.022	96.203	0.000
		36	-0.024	-0.032	98.611	0.000

Datos Semanales

Dependent Variable: R_IGBVL_W
 Method: ARMA Maximum Likelihood (BFGS)
 Date: 10/14/21 Time: 15:41
 Sample: 2 817
 Included observations: 816
 Convergence achieved after 19 iterations
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.176944	0.227021	0.779417	0.4360
AR(1)	0.887378	0.047229	18.78886	0.0000
MA(1)	-0.808535	0.059503	-13.58819	0.0000
SIGMASQ	12.70769	0.249670	50.89792	0.0000
R-squared	0.028233	Mean dependent var		0.163536
Adjusted R-squared	0.024643	S.D. dependent var		3.618416
S.E. of regression	3.573554	Akaike info criterion		5.389987
Sum squared resid	10369.47	Schwarz criterion		5.413048
Log likelihood	-2195.115	Hannan-Quinn criter.		5.398838
F-statistic	7.863856	Durbin-Watson stat		1.941244
Prob(F-statistic)	0.000035			
Inverted AR Roots	.89			
Inverted MA Roots	.81			

Datos Semanales Correlograma del Error Estimado

Date: 10/14/21 Time: 17:18
 Sample: 1 4093
 Included observations: 816

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.029	0.029	0.6703	0.413
		2	0.030	0.029	1.3861	0.500
		3	-0.021	-0.023	1.7536	0.625
		4	-0.064	-0.064	5.1486	0.272
		5	-0.044	-0.039	6.7220	0.242
		6	-0.031	-0.026	7.5274	0.275
		7	-0.067	-0.067	11.282	0.127
		8	0.118	0.119	22.803	0.004
		9	0.055	0.048	25.331	0.003
		10	0.012	-0.005	25.454	0.005
		11	0.047	0.039	27.274	0.004
		12	-0.020	-0.013	27.616	0.006
		13	-0.010	-0.000	27.691	0.010
		14	0.021	0.031	28.047	0.014
		15	-0.012	0.011	28.163	0.021
		16	-0.016	-0.024	28.368	0.029
		17	0.032	0.024	29.224	0.033
		18	-0.041	-0.038	30.599	0.032
		19	0.009	-0.003	30.669	0.044
		20	0.086	0.092	36.922	0.012
		21	-0.031	-0.031	37.718	0.014
		22	-0.024	-0.041	38.211	0.017
		23	-0.001	0.003	38.212	0.024
		24	-0.041	-0.027	39.605	0.024
		25	-0.040	-0.052	40.958	0.023
		26	0.031	0.048	41.776	0.026
		27	-0.019	-0.007	42.067	0.032
		28	-0.013	-0.055	42.217	0.041
		29	0.036	0.036	43.297	0.043
		30	-0.016	-0.009	43.505	0.053
		31	0.001	-0.017	43.507	0.067
		32	-0.071	-0.060	47.779	0.036
		33	-0.050	-0.027	49.929	0.030
		34	0.006	-0.005	49.964	0.038
		35	-0.055	-0.054	52.526	0.029
		36	0.005	0.017	52.546	0.037

Datos Mensuales

Dependent Variable: R_IGBVL_M
 Method: ARMA Maximum Likelihood (BFGS)
 Date: 10/14/21 Time: 15:45
 Sample: 2 189
 Included observations: 188
 Convergence achieved after 3 iterations
 Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.642470	1.017066	0.631690	0.5284
AR(1)	0.126256	0.062925	2.006457	0.0463
AR(2)	0.239294	0.068187	3.509385	0.0006
SIGMASQ	63.92986	3.728383	17.14681	0.0000
R-squared	0.084626	Mean dependent var		0.621668
Adjusted R-squared	0.069702	S.D. dependent var		8.379358
S.E. of regression	8.082057	Akaike info criterion		7.038993
Sum squared resid	12018.81	Schwarz criterion		7.107853
Log likelihood	-657.6653	Hannan-Quinn criter.		7.066892
F-statistic	5.670257	Durbin-Watson stat		2.012547
Prob(F-statistic)	0.000978			
Inverted AR Roots	.56	-.43		

Datos Mensuales Correlograma del Error Estimado

Date: 10/14/21 Time: 17:20
 Sample: 1 4093
 Included observations: 188

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.009	-0.009	0.0149	0.903
		2	0.000	0.000	0.0149	0.993
		3	0.076	0.076	1.1430	0.767
		4	-0.004	-0.003	1.1460	0.887
		5	-0.111	-0.112	3.5590	0.614
		6	0.055	0.048	4.1599	0.655
		7	-0.150	-0.150	8.5766	0.284
		8	-0.057	-0.043	9.2197	0.324
		9	-0.022	-0.032	9.3194	0.408
		10	0.089	0.103	10.912	0.364
		11	0.060	0.083	11.637	0.392
		12	-0.059	-0.094	12.344	0.418
		13	0.021	0.012	12.439	0.492
		14	0.010	-0.027	12.459	0.569
		15	-0.047	-0.030	12.925	0.608
		16	-0.033	-0.044	13.146	0.662
		17	0.029	0.035	13.325	0.714
		18	-0.113	-0.066	16.000	0.593
		19	-0.010	-0.021	16.021	0.656
		20	0.023	0.003	16.130	0.709
		21	-0.038	-0.048	16.436	0.745
		22	0.013	0.027	16.473	0.792
		23	-0.068	-0.112	17.473	0.785
		24	-0.020	-0.007	17.557	0.824
		25	-0.030	-0.051	17.760	0.852
		26	-0.055	-0.059	18.438	0.859
		27	-0.024	-0.022	18.565	0.885
		28	-0.022	-0.048	18.672	0.908
		29	-0.035	0.004	18.947	0.923
		30	0.004	-0.054	18.951	0.941
		31	0.007	-0.004	18.963	0.956
		32	0.031	0.017	19.181	0.964
		33	0.030	0.010	19.393	0.971
		34	0.014	0.005	19.441	0.979
		35	0.099	0.084	21.728	0.961
		36	-0.081	-0.094	23.266	0.950