

¿Cuál es la frecuencia de muestreo óptima en el mercado intradiario del Usd/Cop?

Javier Sandoval

*Profesor/investigador de la Facultad de Finanzas,
Gobierno y Relaciones Internacionales
de la Universidad Externado de Colombia.
E-mail: javier.sandoval@uexternado.edu.co*

*Este documento de trabajo es parte del proyecto de investigación:
Estudio del Mercado Cambiario intradiario colombiano:
Un enfoque microeconómico.*

1. Introducción

El desarrollo de las plataformas transaccionales, ha permitido el estudio cada vez más amplio de los mercados financieros desde la perspectiva intradiaria. Sin embargo, al mismo tiempo que los investigadores han podido ampliar el estado del arte actual de la economía financiera, se han encontrado con los problemas propios de utilizar datos de alta frecuencia. Por esta razón, antes de abordar la implementación de un estudio o aplicación práctica con datos intradiarios, es necesario conocer y controlar la problemática propia de las series.

En general, los trabajos especializados se han enfocado en dos diferentes áreas: estimación de volatilidad realizada¹ y microestructura de mercado. Como muy bien lo explica Andersen et al. 2003, la volatilidad realizada puede considerarse (bajo ciertas condiciones especiales), un estimador ex-post insesgado de la volatilidad diaria de los retornos. Así mismo, este estudio provee al lector evidencia empírica de la superioridad de un modelo vectorial autorregresivo gaussiano, utilizando como entrada la volatilidad realizada diaria. Sin embargo, la esti-

mación de la volatilidad realizada se ve limitada por la aplicación práctica de las condiciones de mercado bajo las cuales estos resultados anteriores son ciertos. Elementos como la ausencia de fricciones de mercado, algunas veces son necesarios para que la volatilidad realizada no pierda su propiedad de estimador insesgado de la variación cuadrática del proceso de los retornos logarítmicos².

La siguiente aplicación de las series intradiarias de precios se ha enfocado en la definición y estudio de la microestructura de mercados financieros. Varios autores, entre ellos Stoll (1978), Glosten and Milgrom (1995), Roll (1984), han desarrollado modelos de comportamiento de agentes de mercado que explican elementos como el flujo de información y la postura de puntas de compra y venta. La importancia principal de las series de alta frecuencia radica en la puesta al descubierto de los patrones del proceso transaccional, los cuales, varían dependiendo de la estructura de mercado. En general la literatura especializada acepta tres fuentes de ruido que se encuentran en variadas combinaciones, según el mercado financiero en particular. Estas fuentes

1 Del inglés *Realized Volatility*.

2 Un enfoque de la literatura especializada consiste en modelar los precios de mercado de los activos en el corto plazo como semi-martingalas, las cuales tienen asociada un proceso de variación cuadrática.

son: la asimetría de información entre los agentes, el efecto del control de inventario y el costo de transacción o de procesamiento de órdenes³.

Al entender la importancia de los datos intradiarios en la economía financiera, este documento tendrá la siguiente estructura. La sección 2 ampliará a las implicaciones teóricas y prácticas de una mala escogencia del espacio de muestreo de los datos intradiarios. La sección 3 desarrollará el modelo de mercado, describirá las características asintóticas del estimador de volatilidad realizada y desarrollará la función de pérdida utilizada para escoger el muestreo óptimo del estimador. La sección 4 incluirá una aplicación para el mercado USD/COP de Colombia. Finalmente, la sección 5 reunirá las conclusiones⁴.

2. Implicaciones de una mala escogencia de la frecuencia de muestreo de los datos intradiarios

Es posible inferir de la teoría estadística que más información es preferible a menos. Por ejemplo, en el

caso particular de la estimación de volatilidad, más información significa aumentar la frecuencia respecto a la cual se toman observaciones de los precios de los activos. De esta forma, un modelo mensual incorporará información omitida en frecuencias anuales y así sucesivamente. Sin embargo, como bien lo menciona Andersen et al. (2003), el enfoque de estudio de la volatilidad diaria sigue en gran medida centrándose en la utilización de retornos diarios, pese a la existencia de información intradiaria amplia y confiable .

En general, existen dos problemas principales cuando se trabaja con series intradiarias. Primero, las series frecuentemente no cuentan con períodos homogéneos de espaciado. Segundo, la existencia de un ruido microeconómico, propio de la estructura de cada mercado, imposibilita la observación limpia de los precios. Aunque el primer problema es considerable y puede llevar a errores profundos en los cálculos, este documento se enfocará en el segundo problema.

¿Por qué si en la práctica existe un ruido microeconómico éste solo afecta a las series intradiarias? Afor-

3 Para introducir al lector en estos conceptos se recomienda Huang and Stoll 1997.

4 Este documento seguirá la línea de los avances encontrados en Bandi and Russell (2005) quien a la vez basa su trabajo en documentos producidos por French et al. (1987), Schwert (1989), Schwert (1990a), Schwert (1990b), Schwert and Seguin (1991), Andersen et al. (2001), Andersen et al. (2003) y Barndor-Nielsen and Shephard (2002).

tunadamente, la respuesta para esta pregunta es sencilla. Como explica Bandi y Russell (2005, 7), el componente de volatilidad de los cambios en los precios logarítmicos de los activos es de orden $O(\sqrt{dt})^5$ mientras que la volatilidad del ruido microeconómico es de orden $O(1)$. En otras palabras, a medida que la frecuencia de muestreo aumenta, las variaciones en el logarítmico de los precios tienden a ser explicados en una mayor cuantía por información que nada tienen que ver con la variabilidad propia del precio eficiente del activo financiero.

Esta información adicional contribuye a la aparición de una varianza espuria en los log-retornos observados que en un modelo como el expuesto en Ait-Sahalia et al. (2003), corresponderá a dos veces la varianza del ruido microeconómico⁶. Afortunadamente, el anterior problema puede ser tratado a través de dos metodológicas las cuales se diferencian por la definición del espaciado óptimo.

Por un lado, Ait-Sahalia et al. (2003) propone modificar la función de máxima verosimilitud del estima-

dor de volatilidad realizada de los retornos logarítmicos para incorporar el ruido microeconómico⁷.

Por otro lado, autores como Bandi and Russell (2005), estiman que no es necesario modificar la función de verosimilitud del estimador de la volatilidad diaria de los retornos logarítmicos sino, por el contrario, modificar el espaciado para que este último optimice su carácter de insesgabilidad y eficiencia. Para lograr este objetivo, los autores proponen un modelo que minimice el error cuadrático medio (MSE por sus siglas en inglés) del estimador de la volatilidad realizada diaria, modificando el espacio entre observaciones (Δ).

El MSE tendrá como variables de entrada adicionales el segundo y cuarto momento del ruido microeconómico y la cuartecidad⁸ de los retornos logarítmicos del activo financiero. En general, se estima que, entre más grande sea el radio del cuadrado de la varianza del ruido respecto a la cuartecidad de los log-retornos eficientes, mayor será Δ . Por lo tanto, el espaciado óptimo que sugiere Ait-Sahalia et

5 Para entender este concepto se recomienda al lector verificar el modelo de comportamiento de activos descrito en la sección 3.

6 Para mayores detalles sobre esta afirmación consultar Ait-Sahalia et al. 2003, p. 359.

7 Detalle de las variaciones en la función de máxima verosimilitud del estimador se pueden observar en Ait-Sahalia et al. 2003, pp. 367.

8 Para una introducción al concepto de cuartecidad o variación cuártica se recomienda al lector Barndor, Nielsen y Shephard (2002).

al. 2003 será tan pequeño como sea posible mientras que el que sugiere Bandi and Russell 2005 se ubicará en algún punto entre el mínimo posible y el período correspondiente a un día.

En lo que concierne al documento, el espaciado óptimo se enfocará en el problema de modelar la volatilidad diaria a partir de datos intradiarios siguiendo la metodología de Bandi and Russell (2005).

3. Modelo

3.1. Supuestos del Modelo

El proceso del precio eficiente del activo financiero y el ruido macroeconómico seguirán los siguientes supuestos,

- los log-precios no contaminados estarán modelados por un proceso estocástico continuo conformado por un proceso de variación acotada y una martingala local definidos en el intervalo $[0; T]$. De esta forma, el proceso de los log-precios limpios será una semi-martingala.

$$\log(p_t) = A(t) + \int_0^t \sigma_s dw_s \quad 0 \leq t \leq T \quad (1)$$

$w(t)$ será un movimiento browniano unidimensional.

- Se asumirá que la función $\sigma(t)$ será cuadrado integrable.

$$\int_0^t \sigma_s^2 ds < \infty \quad (2)$$

- El proceso de los log-precios sólo estará gobernado por la parte estocástica, Así, $A(t) = 0$. Este supuesto es válido debido a que en instantes de tiempo corto ignorar $A(t)$ no causa distorsiones significativas.
- σ_t , será un proceso RCLL (continuo por derecha con límites por izquierda) acotado e independiente de w_t .

3.2. Precios y retornos contaminados

Esta sección sigue la notación y estructura en Barndor-Nielsen and Shephard (2002) y Bandi and Russell (2005).

Primero se definirá una partición uniformemente espaciada sobre el intervalo de tiempo $[0, T]$. Así, $\{0 = t_0 < t_1 < \dots < t_i < t_N = T\}$. Luego, se definirán nuevas particiones para cada uno de los intervalos de tiempo $t_i - t_{i-1}$ como $\{t_{i-1} = t_{i-1,0} < t_{i-1,1} < t_{i-1,j} < \dots < t_{i-1,M} = t_i\}$. Para simplificar la notación, i, j se deberá entender de ahora en adelante como $t_{i,j}$.

Sobre estas particiones se definirá el siguiente proceso discreto,

$$\tilde{p}_{i,j} = p_{i,j} \eta_{i,j} \quad (3)$$

donde $p_{i,j}$ se interpretará como el precio del activo en ausencia de contaminación y $\eta_{i,j}$ será el ruido microeconómico ambos observados en el instante $t_{i,j}$.

Intuitivamente, el precio contaminado del activo financiero, $\tilde{p}_{i,j}$, es observado en M instantes uniformemente espaciados de tiempo, todos ubicados dentro de un día específico i . Es importante resaltar que definiendo Δt_i ,

como un día transable, dependiendo del M escogido, habrá que interpretar apropiadamente la longitud del subespaciado. Así, si $M = 5$ o $M = 150$ el espaciado deberá entenderse como 1 hora o 5 minutos respectivamente⁹.

Partiendo de la ecuación (3), se deduce que los log-retornos serán de la forma,

$$\underbrace{\log(\tilde{p}_{i,j}) - \log(\tilde{p}_{i,(j-1)})}_{r_{i,j}} = \underbrace{(\log(p_{i,j}) - \log(p_{i,(j-1)}))}_{r_{i,j}} + \underbrace{(\log(\eta_{i,j}) - \log(\eta_{i,(j-1)}))}_{\in_{i,j}} \quad (4)$$

Así, los retornos contaminados $\tilde{r}_{i,j}$ estarán conformados por dos componentes. Los retornos del precio eficiente, $r_{i,j}$, los cuales, representan genuinos cambios en el precio del activo financiero y el ruido microeconómico, $\in_{i,j}$, el cual, puede ser referenciado a algunas de las causas mencionadas en la sección 1.

Una vez definido el proceso de precios contaminados, algunos supuestos adicionales serán,

- la contaminación de los log-precios $\log(\eta_{i,j})$, la cual se denotará de ahora en adelante $\tilde{\eta}_{i,j}$, será iid con media cero y acotada hasta el cuarto momento.

$$\blacksquare E(r_{i,j}^2) = E\left[\left(\int_{i,(j-1)}^{i,j} \sigma_s dw_s\right)^2\right] = \int_{i,(j-1)}^{i,j} \sigma_s^2 ds$$

debido a la isometría de $It\hat{o}$.

Por último, $\tilde{\eta}_{i,j}$ será independiente de $r_{i,j}$ para todo i, j .

3.3. Presencia de autocorrelación en el proceso de los retornos de los activos financieros

A pesar que los retornos eficientes $r_{i,j}$ se asumen independientes del ruido

9 Esta interpretación se toma para un día transable de cinco horas.

microeconómico $\eta_{i,j}$, la serie de retornos contaminados $\tilde{r}_{i,j}$, $i, j = 1, 2$, incorporará autocorrelaciones de orden 1.

Esto se explica debido a que $\epsilon_{i,j}$ estará correlacionado negativamente con $\epsilon_{i,(j-1)}$.

$$\begin{aligned}
 COV(\tilde{r}_{i,j}, \tilde{r}_{i,(j-1)}) &= COV(r_{i,j} + \epsilon_{i,j}, r_{i,j} + \epsilon_{i,(j-1)}) \\
 &= E[(r_{i,j} + \epsilon_{i,j}) \times (r_{i,j} + \epsilon_{i,(j-1)})] \\
 &= E[\epsilon_{i,j}\epsilon_{i,(j-1)}] \text{ , debido a la independencia de } r_t \quad (5) \\
 &= E[(\tilde{\eta}_{i,j} - \tilde{\eta}_{i,(j-1)}) \times (\tilde{\eta}_{i,(j-1)} - \tilde{\eta}_{i,(j-2)})] \\
 &= -E[\tilde{\eta}_{i,(j-1)}^2] \\
 &= -\sigma_\eta^2
 \end{aligned}$$

Este hecho teórico ha sido ampliamente documentado en la práctica y resalta en mayor medida en las series de precios de cierre debido a la mayor varianza del ruido microeconómico (Ver Figura 3).

3.4. Características del estimador de volatilidad realizada en presencia de ruido microeconómico

Al igual que en Andersen et al. (2003), dentro del contexto del anterior modelo un estimador natural para la volatilidad diaria podrá ser

$$\hat{V}_i = \sum_{j=1}^M r_{i,j}^2 \quad (6)$$

Esto se debe a que con una partición cada vez más fina de retornos cuadráticos intradiarios (limpios) para el día

i , la serie converge a la volatilidad realizada del día en particular. Así,

$$\lim_{M \rightarrow \infty} \hat{V}_i = \int_{(i-1)}^i \sigma_s^2 ds = \sigma^{*2}(i). \quad (7)$$

Este resultado se deriva directamente de los supuestos planteados en la sección 3.1 (la anterior convergencia se debe entender en probabilidad).

Los retornos limpios generalmente, nunca pueden ser observados directamente en el mercado. Por lo tanto, se podrá plantear como una generalización del anterior caso, otro estimador para la volatilidad realizada diaria para el día i . Así ,

$$\tilde{V}_i = \sum_{j=1}^M \tilde{r}_{i,j}^2 \quad (8)$$

Sin embargo, nada garantiza que la serie de retornos contaminados sea acotada y más aún, que converja ha-

cia la volatilidad realizada cuando M tiende a infinito. Por esta razón, el primer elemento de estudio será encontrar las propiedades asintóticas de \tilde{V}_i cuando M tiende a infinito.

Recordando la ecuación 4, el retorno contaminado de los activos financieros estará dado por

$$\tilde{r}_{i,j} = r_{i,j} + \epsilon_{i,j}$$

De esta forma, estudiar las propiedades asintóticas de la serie de retornos cuadráticos contaminados, equivale a estudiar las propiedades asintóticas de las series en que puede descomponerse

$$\sum_{j=1}^M r_{i,j}^2$$

para cada día i .

Así, se estudiará qué sucede individualmente con los términos A, B y C,

$$\tilde{V}_i = \underbrace{\sum_{j=1}^M r_{i,j}^2}_A + \underbrace{\sum_{j=1}^M \tilde{\epsilon}_{i,j}^2}_B + 2 \underbrace{\sum_{j=1}^M r_{i,j} \tilde{\epsilon}_{i,j}}_C \quad (9)$$

cuando $M \rightarrow \infty$.

Proposición 1. Cuando $M \rightarrow \infty$, el término A tiene como distribución asintótica,

$$\sqrt{M}(A_i - V_i) \xrightarrow{M \rightarrow \infty} N(0, 2Q_i) \quad (10)$$

Condicional a la realización de σ_i .

Donde, Q_i representa la variación cuártica de los log-precios limpios

$$\int_{i-1}^i \sigma_s^4 ds$$

y V_i la variación cuadrática. Además, como lo muestra Chung y Williams (1990) por ejemplo, el término A converge (en probabilidad) a la variación cuadrática del proceso en i , o en otras palabras, a la volatilidad diaria definida como se hizo en la ecuación 7.

Demostración. Ver apéndice en Bandi y Russell (2002).

Proposición 2. Cuando $M \rightarrow \infty$, el término B tiene como características asintóticas,

$$\sqrt{M} \left(\frac{B_i}{M} - E[\epsilon^2] \right) \xrightarrow{M \rightarrow \infty} N(0, 4E[\eta^4]) \quad (11)$$

La ecuación 11 implica que el término B explota c.s. a medida que crece M .

La explicación intuitiva para esto consiste en entender que lo que se acumula en el término

$$\sum_1^M \tilde{\epsilon}_{i,j}$$

es ruido microeconómico que gana más peso entre más fina sea la partición de muestreo.

Demostración. Ver apéndice en Bandi y Russell (2002).

Proposición 3. Cuando $M \rightarrow \infty$, el término C tiene como características asintóticas,

$$C_i \xrightarrow{M \rightarrow \infty} N(0, 2\sigma_\eta^2 V_i)$$

Condicional a la realización de σ_t . (12)

La ecuación 12 intuitivamente significa que a pesar de que se esté acumulando ruido, este último es multiplicado por un término cada vez más pequeño a medida que crece M . La anterior convergencia es en sentido débil.

Demostración. Ver apéndice Bandi y Russell (2002).

Proposición 4. las anteriores proposiciones implican que el término V_i explota cuando M tiende a infinito,

$$A_i + B_i + C_i = \left(V_i + O_p\left(\frac{1}{\sqrt{M}}\right) \right) + \left(ME(\epsilon^2) + O_p\left(\sqrt{M}\right) \right) + O_p(1) \xrightarrow[M \rightarrow \infty]{c.s.} \infty^{10} \quad (13)$$

Como conclusión de las anteriores 4 proposiciones es posible mostrar que el estimador \tilde{V}_i de la volatilidad diaria explota c.s. a medida que la partición se hace más fina o que M se hace más grande. Todo este efecto es liderado por la acumulación de ruido del término B , la cual, se hace más crítica

cuanto más información agregamos al cálculo de \tilde{V}_i . Por lo tanto, es necesario encontrar el intervalo de muestreo óptimo que maximice el uso de la información versus el sesgo que esta última infiere en los cálculos.

3.5. Estimación y minimización del error cuadrático medio condicional de V_i

Basados en los hallazgos de la sección anterior, es necesario desarrollar una aproximación sistemática al problema de la acumulación de errores. Para ello se plantea como en Bandi y Russell (2002), la escogencia de la partición del espacio muestral buscando minimizar el error cuadrático medio condicional.

Intuitivamente, lo que se buscará será ponderar en la función de MSE lo que se gana en términos de eficiencia contra lo que se pierde en términos de sesgo del estimador V_i por agregar una partición más fina de los retornos contaminados.

El error cuadrático medio del estimador de la volatilidad realizada para el día i estará dado por,

$$E_\sigma \left[\tilde{V}_i - \int_{i-1}^i \sigma_s^2 ds \right]^2 = 2 \frac{1}{M} (Q_i + z) + M\beta + M^2\alpha + \gamma \quad \text{con } z \xrightarrow{c.s.} 0 \quad (14)$$

10 Entiéndase $O_p(1)$ como acotado en probabilidad.

donde los parámetros α , β y γ estarán dados por,

$$\begin{aligned}\alpha &= E[\epsilon^2]^2 \\ \beta &= 2E[\epsilon^4] - 3(E[\epsilon^2])^2 \\ \gamma &= 4E[\epsilon^2]V_i - 2E[\epsilon^4] + 2(E[\epsilon^2])^2\end{aligned}\quad (15)$$

Demostración. Ver apéndice.

Algunas observaciones relevantes sobre la ecuación 14 son las siguientes.

– En ausencia de ruido microeconómico el MSE del estimador V_i se reduce a la varianza del término A expuesta en la ecuación 10.

– Debido al ruido microeconómico, el MSE explota a medida que M crece. Por lo tanto, recordando que el MSE de un estimador está constituido por la varianza del estimador y el sesgo relativo, lo que se observa es un sesgo creciente a medida que más información es incorporada.

– Minimizando el MSE con respecto a M se podrá encontrar el número de particiones óptimas con respecto al grado de insesgadez y eficiencia del estimador \tilde{V}_i .

Basados en la ecuación 14 podemos definir el número de particiones óptimas en el muestreo de una serie de retornos contaminados como el valor M^* que cumple la siguiente condición,

$$\begin{aligned}\{M^* := \operatorname{argmin}(2\frac{1}{M}Q_i + M\beta + M^2\alpha + \gamma)\} \text{ o,} \\ \{M^* := M \mid -2\frac{1}{M^2}Q_i + \beta + M\alpha = 0\}\end{aligned}\quad (16)$$

Resolviendo la ecuación planteada en 16 podremos encontrar el espaciado adecuado para realizar muestreos de series de retornos contaminados. Hay que recordar que cada serie de retornos de activos tendrá sus propias características de ruido y, por lo tanto, su propio espaciado óptimo. Sin embargo, intuitivamente, es posible decir que entre más alta sea la variabilidad del activo respecto a la variabilidad de la contaminación microeconómica, el espaciado podrá ser más fino.

3.5.1. Estimadores de la variación cuártica y el segundo y cuarto momento del ruido microeconómico

La sección anterior plantea la necesidad de conocer la variación cuártica del proceso de log-precios y el segundo y cuarto momento del ruido microeconómico para poder realizar el cálculo de minimización del MSE. Sin embargo, éstos son parámetros desconocidos para el modelador y, por lo tanto, deben ser estimados.

En ausencia de ruido microeconómico, Barndor-Nielsen and Shephard (2002) presenta un estimador con las características deseables para la variación cuártica. Sin embargo, la contaminación de la serie de log-precios hace que el estimador sea ineficiente y tenga características asintóticas no deseables.

Ante ese problema, Bandi and Russell (2005) propone desarrollar un muestreo a bajas frecuencias del estimador para poder tener resultados aceptables. El efecto de la escogencia del muestreo se considera bajo en la estimación del MSE condicional, aunque los autores sugieren un espaciado de 15 min. para activos líquidos según resultados obtenidos bajo simulaciones.

Proposición 5. Siguiendo el anterior análisis, el estimador de la variación cuártica para el día i será,

$$\frac{M}{3} \sum_{j=1}^M \tilde{r}_{i,j}^4 \quad (17)$$

Demostración. Ver Barndor-Nielsen and Shephard (2002) para la prueba en ausencia de ruido microeconómico y Bandi and Russell (2002) para los resultados de estabilidad por simulaciones.

Proposición 6. El cuarto momento del ruido microeconómico puede ser aproximado usando el estimador,

$$\frac{1}{M} \sum_{j=1}^M \tilde{r}_{i,j}^4 \rightarrow E[\epsilon^4] \quad (18)$$

M debe ser escogida tan grande como sea posible.

Demostración. Ver apéndice a de Bandi and Russell (2002). Esta convergencia se garantiza en probabilidad bajo los supuestos iniciales.

Proposición 7. Un promedio simple de los retornos cuadráticos contaminados analizados al espaciado más pequeño posible, estima consistentemente el segundo momento del ruido microeconómico.

$$\frac{1}{M} \sum_{j=1}^M \tilde{r}_{i,j}^2 \rightarrow E[\epsilon^2] \quad (19)$$

Demostración. Ver teorema 1 de Bandi and Russell (2002). Esta convergencia también se debe leer en probabilidad.

4. Aplicación del modelo al mercado USD/COP intradiario colombiano

Los datos utilizados para estimar el espaciado óptimo comprende la serie de puntos medios de las puntas de compra y venta del USD/COP para 257 días transables comenzando el 1o. de octubre de 2004. El punto medio se ha utilizado a diferencia de los precios de cierre debido a dos problemas observados con estos últimos. Primero, no se tienen datos satisfactorios para espaciados menores a 15 minutos. Además, se desea eliminar gran parte del ruido microeconómico que en general corresponderá a saltos del precio producidos por el inicio de operaciones tanto desde el lado de la compra como de la venta. La Figura 1 muestra la evolución del tipo de cambio para el período estudiado.

La serie de puntos medios por su parte, no es una serie uniformemente espaciada. Para corregir este inconveniente, se creó una serie sintética uniforme escogiendo las observaciones en los momentos del tiempo que indicaba el intervalo utilizado. Sin embargo, cuando no se encontró dato alguno, se utilizó el más cercano disponible.

Algunas estadísticas de la serie de puntos medios para intervalos de 30 minutos se muestran en la Tabla 1. De los datos se pueden resaltar las siguientes observaciones:

- El número de pares de puntas de compra y venta colocados durante los 257 días transables fue de 242; 232.

- Los datos analizados presentan una tendencia negativa con una tasa de devaluación para el período de -13.73 %c.c.

- La volatilidad media diaria para el período analizado, se ubicó en 0.32% (calculada a partir del espaciado óptimo)

- El retorno max y min. son respectivamente, 1.575% y -1.4% (medido en un espaciado de 30 min.).

- El spread medio se ubicó en 1.76 pesos.

4.1. Cálculo del espaciado óptimo

Aproximando los momentos del ruido microeconómico utilizando un espaciado de 1 min. y la variación cuártica

de la volatilidad con un espaciado de 20 minutos, los resultados pueden observarse en el Cuadro 2.

Por un lado, se encuentra que el espaciado óptimo se ubica cercano a los 10 minutos entre observaciones. Por lo tanto, dentro del contexto del modelo planteado, cualquier intervalo menor a 10 minutos, utilizado para cálculos econométricos, implica estar expuesto al problema del ruido microeconómico.

Los problemas en los que se incurre cuando no se optimiza el intervalo de muestreo, pueden ser fácilmente observado en los siguientes ejemplos.

La Figura 4 muestra el error estándar medio del estimador de volatilidad realizada diaria en función del espaciado de las observaciones. Como es claro, después de la barrera de los 11 minutos, el error medio cuadrático comienza a mostrar un crecimiento considerable. Esta diferencia observada, llevará a que el estimador de volatilidad diaria explote a medida que la frecuencia de muestreo converja a cero.

Sin embargo, el escenario anterior no es el único donde se pueden observar los efectos de la acumulación de ruido en los estimadores de volatilidad. La Figura 5 muestra la serie de volatilidad realizada diaria para el período en observación. La línea roja muestra el cálculo desarrollado para un intervalo de 5 minutos. La línea

azul utiliza el intervalo óptimo. Como se puede observar, la línea roja por lo general sobrepasa la línea azul mostrando siempre una sobreestimación de la volatilidad, que en algunos lugares llega a ser del 55 %. La sobreestimación utilizando un espaciado de 5 minutos es observada mejor en la Figura 6. En general, es posible destacar que la sobreestimación se produce en mayor medida cuando la volatilidad aumenta.

5. Conclusiones

El grado de sofisticación de los mercados financieros ha producido la proliferación de datos intradiarios para activos de alta liquidez. Sin embargo, a pesar de la disponibilidad de estos datos, los modelos de estimación de volatilidad siguen, en gran medida, calibrándose con observaciones diarias. Las principales razones que evitan el incremento de la importancia de los datos intradiarios en los modelos de volatilidad son principalmente, problemas de espaciados no homogé-

neos y exposición al ruido microeconómico. Este artículo ha expuesto una posible metodología de aproximación al segundo problema, trabajando sobre la base de avances alcanzados por Bandi and Russell (2002) y Barndor-Nielsen and Shephard (2002) y aplicándola directamente al mercado del USD/COP.

Sin embargo, no hay que olvidar que los hallazgos encontrados se fundamentan sobre argumentos de comportamiento asintótico y diferentes tipos de convergencias de las variables discretizadas del modelo.

A pesar de esta debilidad, es claro que ignorar la presencia del ruido microeconómico puede llevar a grandes sesgos en la estimación de la volatilidad realizada del USD/COP y, al final, al fracaso de cualquier modelo construido sobre estos resultados.

Por último, teniendo en cuenta las consideraciones anteriores, este documento ha mostrado que el muestreo óptimo intradiario en la serie del USD/COP debe realizarse en espaciados del orden de 10 minutos.

A. Gráficas

Figura 1

Evolución del tipo de cambio USD/COP durante el período estudiado

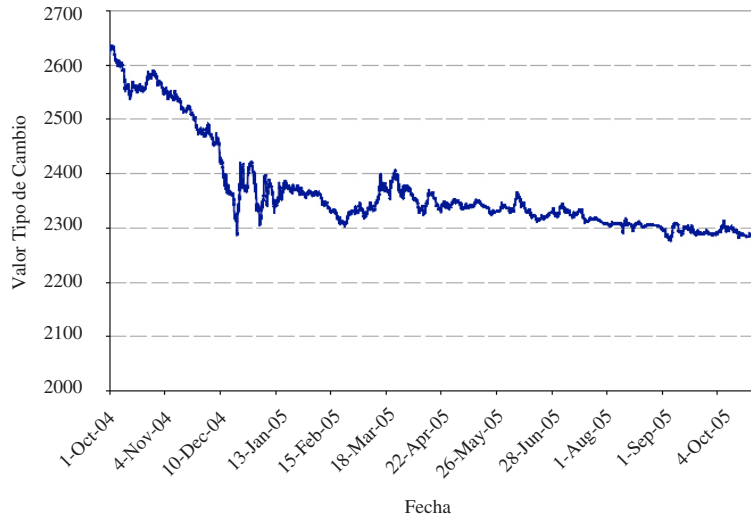


Figura 2

Evolución de la serie de retornos del USD/COP durante el período estudiado

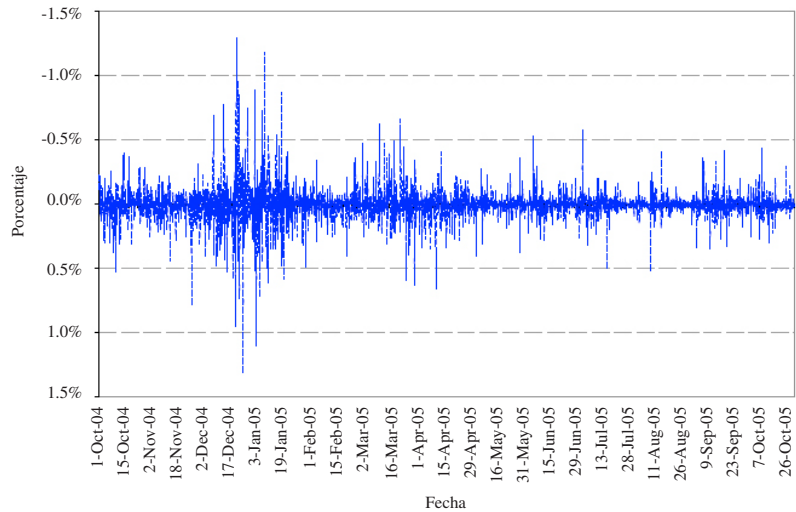


Figura 3

Autocorrelaciones de la serie precios de cierre
espaciados cada 10 minutos

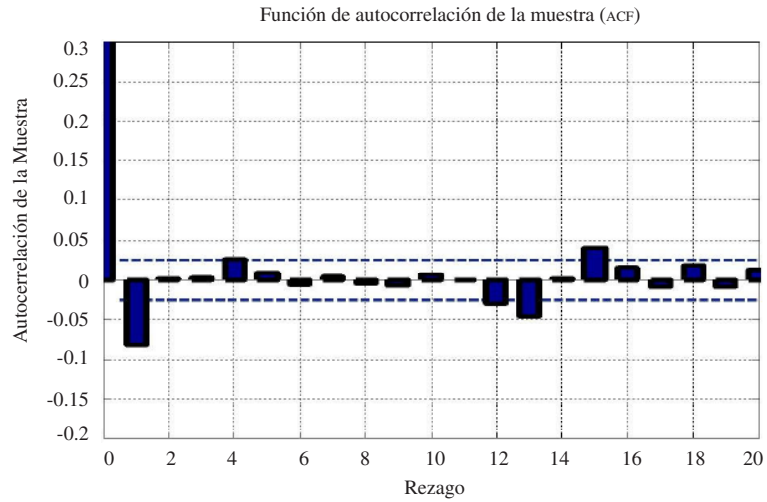


Figura 4

MSE del estimador de la volatilidad realizada diaria
en función del espaciado utilizado (en minutos)

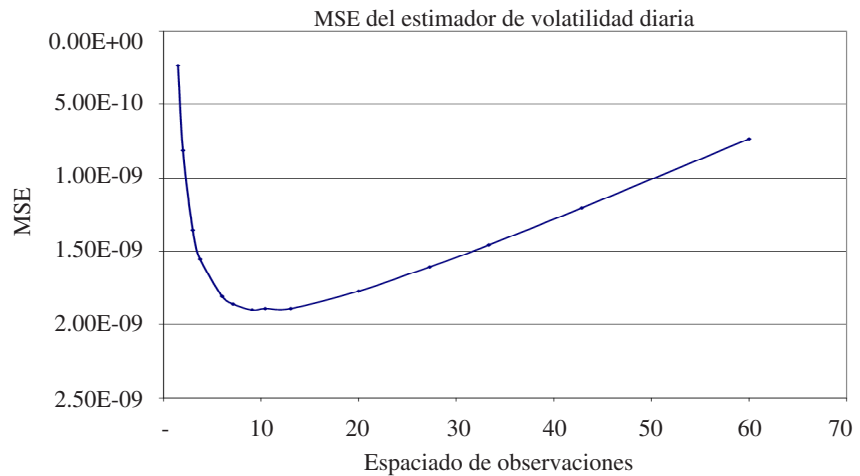


Figura 5

Volatilidad realizada diaria calculada a partir de espaciados de 10 y 5 minutos

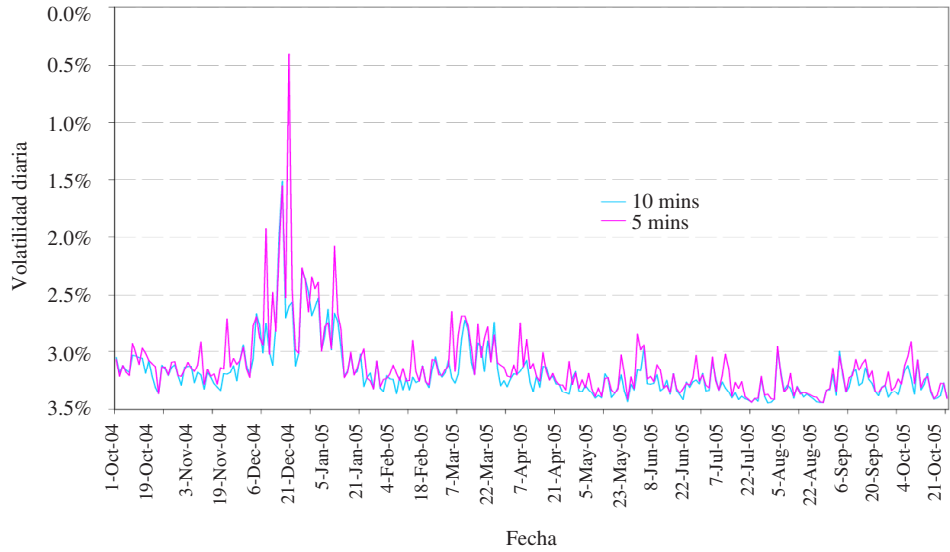
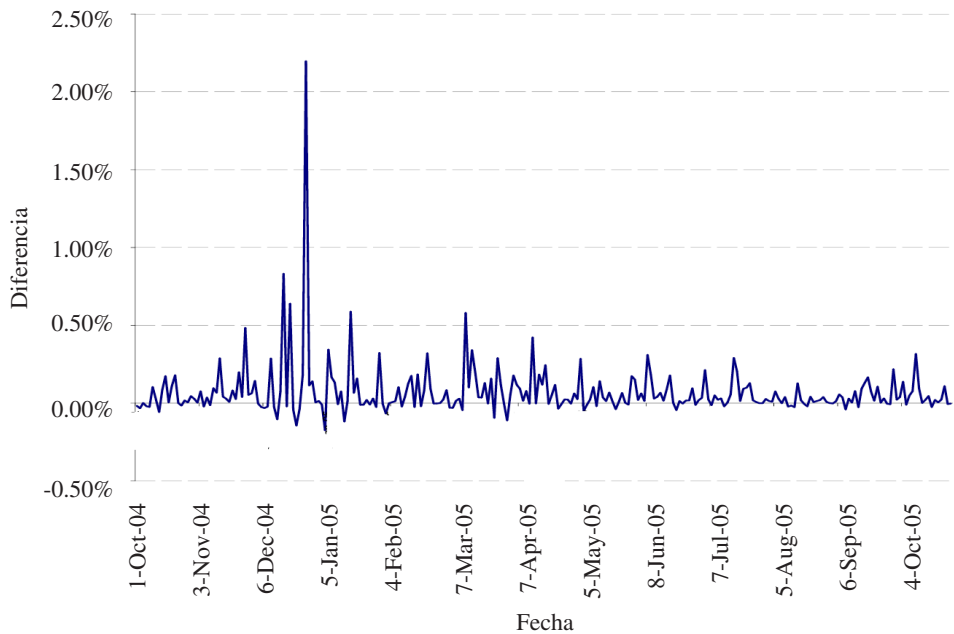


Figura 6

Diferencia entre la volatilidad realizada diaria calculada a partir de espaciados de 10 y 5 minutos



B. Tablas

Cuadro 1
Estadísticas de los datos de la muestra

	Precios	Retornos
Mínimo	2276.93	-1.399%
Máximo	2,636.58	1.575%
Media	2,368.03	-0.005%
Mediana	2,338.40	-0.003%
Desvest	84.84	0.155%
Varianza	7,198.62	2.402E-06
Coef de asimetría	1.55	0.362
Kurtosis	4.26	19.4623

Cuadro 2
Resultados de la minimización del MSE del estimador
de la volatilidad realizada promedio para el período en estudio

Días de la serie	258
Datos por día	20 min 15
Datos por día	1 min 300
Aproximación cuarticidad	30 min 4.281E-09
Aproximación segundo momento ruido	7.386E-08
Aproximación al cuarto momento ruido	5.025E-12
Vi	2.166E-05
a (alpha)	5.455E-15
(beta)	1.003E-11
Cuarticidad	4.281E-09
M*	29
M* (en minutos)	10.42813885
Derivada de la función del MSE	0
MSE	5.922E-10

C. Anexos

Derivación de las ecuaciones 14 y 15.

Expandiendo la parte derecha de la ecuación 14 se tendrá,

Demostración.

$$\begin{aligned}
& E_{\sigma} \left[\sum_{j=1}^M (r_{(i,j)} + \epsilon_{(i,j)})^2 - \int_{i-1}^i \sigma_s^2 ds \right]^2 \\
&= E_{\sigma} \left[\sum_{j=1}^M (r_{(i,j)}^2 + \epsilon_{(i,j)}^2 + 2r_{(i,j)}\epsilon_{(i,j)}) - \int_{i-1}^i \sigma_s^2 ds \right]^2 \\
&= E_{\sigma} \left[\underbrace{\sum_{j=1}^M r_{(i,j)}^2 - \int_{i-1}^i \sigma_s^2 ds}_A \right]^2 + E_{\sigma} \left[\underbrace{\sum_{j=1}^M (\epsilon_{(i,j)}^2 + 2r_{(i,j)}\epsilon_{(i,j)})}_B \right]^2 + \quad (20) \\
&\quad \underbrace{2E_{\sigma} \left[\left(\sum_{j=1}^M r_{(i,j)}^2 - \int_{i-1}^i \sigma_s^2 ds \right) \left(\sum_{j=1}^M (\epsilon_{(i,j)}^2 + 2r_{(i,j)}\epsilon_{(i,j)}) \right) \right]}_C
\end{aligned}$$

Ahora se analizará término por término. Primero el término B.

$$\begin{aligned}
& E_{\sigma} \left[\sum_{j=1}^M (\epsilon_{i,j}^2 + 2r_{i,j}\epsilon_{i,j}) \right]^2 \\
&= E_{\sigma} \left[\sum_{j=1}^M \sum_{g=1}^M (\epsilon_{i,j}^2 + 2r_{i,j}\epsilon_{i,j}) (\epsilon_{i,g}^2 + 2r_{i,g}\epsilon_{i,g}) \right] \\
&= E_{\sigma} \left[\sum_{j=1}^M \sum_{g=1}^M (\epsilon_{i,j}^2 \epsilon_{i,g}^2) + 2 \sum_{j=1}^M \sum_{g=1}^M (\epsilon_{i,j}^2 r_{i,g} \epsilon_{i,g}) + 2 \sum_{j=1}^M \sum_{g=1}^M (\epsilon_{i,j} r_{i,g} \epsilon_{i,g}^2) + 4 \sum_{j=1}^M \sum_{g=1}^M (\epsilon_{i,j} r_{i,g} r_{i,j} \epsilon_{i,g}) \right] \\
&= E_{\sigma} \left[\sum_{j=1}^M \sum_{g=1}^M (\epsilon_{i,j}^2 \epsilon_{i,g}^2) \right] + 4E_{\sigma} \left[\sum_{j=1}^M \sum_{g=1}^M (\epsilon_{i,j} r_{i,g} r_{i,j} \epsilon_{i,g}) \right] \\
&= \sum_{j=1}^M E(\epsilon_{i,j}^4) + 2 \sum_{j=1}^M \sum_{j < g} E(\epsilon_{i,j}^2 \epsilon_{i,g}^2) + \\
&\quad 4 \left(\sum_{j=1}^M E_{\sigma}(r_{i,j}^2 \epsilon_{i,j}^2) \right) + 8 \sum_{j=1}^M \sum_{j < g} E_{\sigma}(\epsilon_{i,j} r_{i,g} r_{i,j} \epsilon_{i,g})
\end{aligned} \quad (21)$$

$$\begin{aligned}
&=ME(\epsilon^4) + 2(M-1)E(\epsilon_i^2\epsilon_{-1,i}^2) + (M^2-3M+2)(E(\epsilon^2))^2 + 4E(\epsilon^2)V_i \\
&=M^2(E(\epsilon^2))^2 + M(E(\epsilon^4) + 2\epsilon_i^2\epsilon_{-1,i}^2) - 3(E(\epsilon^2))^2 + 4E(\epsilon^2)V_i - 2E(\epsilon_i^2\epsilon_{-1,i}^2) + 2(E(\epsilon^2))^2 \\
&=M^2\alpha + M\beta + \gamma
\end{aligned}$$

donde,

$$\begin{aligned}
\alpha &= (E(\epsilon^2))^2 \\
\beta &= E(\epsilon^4) + 2E(\epsilon_i^2\epsilon_{-1,i}^2) - 3E(\epsilon^2)^2 \\
\gamma &= 4E(\epsilon^2)V_i - 2E(\epsilon_i^2\epsilon_{-1,i}^2) + 2(E(\epsilon^2))^2 \\
&Y, \\
E(\epsilon_i^2\epsilon_{-1,i}^2) &= E(\epsilon_i^4)
\end{aligned} \tag{22}$$

Ahora el término C,

$$\begin{aligned}
&2E_\sigma \left[\left(\sum_{j=1}^M r_{(i,j)}^2 - \int_{i-1}^i \sigma_s^2 ds \right) \left(\sum_{j=1}^M (\epsilon_{i,j}^2 + 2r_{i,j}\epsilon_{i,j}) \right) \right] \\
&= E_\sigma \left(\sum_{j=1}^M r_{i,j}^2 \right) E \left(\sum_{g=1}^M \epsilon_{i,g}^2 \right) + 4E_\sigma \left(\sum_{j=1}^M \sum_{g=1}^M E_\sigma(\epsilon_{i,g}r_{i,g}r_{i,j}^2) \right) - 2V_i E_\sigma \left(\sum_{j=1}^M (\epsilon_{i,j}^2 + 2r_{i,j}\epsilon_{i,j}) \right) \\
&= 2MV_i E(\epsilon^2) - 2MV_i E(\epsilon^2) = 0
\end{aligned} \tag{23}$$

Por último el término A,

$$\begin{aligned}
&E_\sigma \left[\sum_{j=1}^M r_{i,j}^2 - \int_{i-1}^i \sigma_s^2 ds \right]^2 \\
&= E_\sigma \left[\sum_{j=1}^M r_{i,j}^2 \right]^2 - 2E_\sigma \left(\sum_{j=1}^M r_{i,j}^2 \right) \left(\int_{i-1}^i \sigma_s^2 ds \right) + \left(\int_{i-1}^i \sigma_s^2 ds \right)^2 \\
&= E_\sigma \left[\sum_{j=1}^M \sum_{g=1}^M (r_{i,j}^2 r_{i,g}^2) \right] - \left(\int_{i-1}^i \sigma_s^2 ds \right)^2 \\
&= E_\sigma \left(\sum_{j=1}^M r_{i,j}^4 \right) + 2 \sum_{j=1}^M \sum_{j < g}^M E_\sigma(r_{i,j}^2 r_{i,g}^2) - \left(\int_{i-1}^i \sigma_s^2 ds \right)^2 \\
&= E_\sigma \left(\sum_{j=1}^M r_{i,j}^4 \right) + 2 \sum_{j=1}^M \sum_{j < g}^M \left(\int_{i,j-1}^{i,j} \sigma_s^2 ds \right) \left(\int_{i,g-1}^{i,g} \sigma_s^2 ds \right) - \left(\int_{i-1}^i \sigma_s^2 ds \right)^2
\end{aligned} \tag{24}$$

$$\begin{aligned}
&= \sum_{j=1}^M E_{\sigma}(r_{i,j}^4) - \sum_{j=1}^M \left(\int_{i,j-1}^{i,j} \sigma_s^2 ds \right)^2 \\
&= \sum_{j=1}^M V_{\sigma}(r_{i,j}^2) + \sum_{j=1}^M \left(E_{\sigma} r_{i,j}^2 \right)^2 - \sum_{j=1}^M \left(\int_{i,j-1}^{i,j} \sigma_s^2 ds \right)^2 \\
&= \sum_{j=1}^M V_{\sigma}(r_{i,j}^2) \\
&\text{como } r_{i,j} \sim N\left(0, \int_{i,j-1}^{i,j} \sigma^2(s) ds\right) \longrightarrow \frac{r_{i,j}^2}{\int_{i,j-1}^{i,j} \sigma^2(s) ds} \sim \chi_1^2 \text{ por lo tanto,} \quad (25) \\
&= 2 \sum_{j=1}^M \left(\int_{i,j-1}^{i,j} \sigma^2(s) ds \right)^2 \\
&= 2 \frac{1}{M} \sum_{j=1}^M M \left(\int_{i,j-1}^{i,j} \sigma^2(s) ds \right)^2 \\
&= 2 \frac{1}{M} (Q_i + z) \text{ donde, } z \xrightarrow{c.s.} 0
\end{aligned}$$

La prueba de esta última convergencia puede encontrarse en el apéndice A de Barndor-Nielsen and Shephard (2002).

Referencias

- Ait-Sahalia, Yacine, and Mykland. (2003). "How often to sample a continuous-time process in the presence of market microstructure noise". Technical report, NBER Working Paper No. W9611.
- Andersen, Bollerslev, and Diebold. (2001). "The distribution of realized stock return volatility". *Journal of Financial Economics*, 61:43{76.
- Andersen, Bollerslev, and Diebold. (2003). "Modeling and forecasting realized volatility". *Econometrica*, 71(2):576{625.
- Bandi and Russell. (2002). "Microstructure noise, realized volatility and optimal sampling". December 22.
- Bandi and Russell. (2005). "Separating microstructure noise from volatil-

- ity”. Technical report, AFA Philadelphia Meetings.
- Barndor-Nielsen and Shephard. (2002). “Econometric analysis of realized volatility and its use in estimating stochastic volatility models”. *Journal of the Royal Statistical Society*, 64:253 {280.
- Chung and Williams. (1990). “Introduction to Stochastic Integration”. Birkhäuser. French, Schwert, and Stambaugh. (1987). “Expected stock returns and volatility”. *Journal of Financial Economics*, 19:3 {29.
- Glosten and Milgrom. (1995). “Bid, ask and transaction prices in a specialist market with heterogeneously informed traders”. *Journal of Financial Economics*, 14(1):71 {100.
- Huang and Stoll. (1997). “The components of the bid-ask spread: A general approach”. *The Review of Financial Studies*, 10(4):995 {1034.
- Roll. (1984). “A simple implicit measure of the effective bid-ask spread in ancient market”. *The Journal of Finance*, 39(4):1127 {1139.
- Schwert. (1989). “Why does stock market volatility change over time?”. *Journal of Finance*, 44:1115 {1153.
- Schwert. (1990a). “Stock market volatility”. *Financial Analysts Journal*, 44:5.
- Schwert. (1990b). “Stock volatility and the crash of the 87’s”. *Review of Financial Studies*, 3:77 {102.
- Schwert and Seguin. (1991). “Heteroskedasticity in stock returns”. *Journal of Finance*, 45:1129 {1155.
- Stoll. (1978). “The supply of dealer services in securities markets”. *The Journal of Finance*, 33(4):1133 {1151.

Javier Sandoval.

“¿Cuál es la frecuencia de muestreo óptima en el mercado intradiario del Usd/Cop?”, en *Odeon*, 2007-2008, núm. 4, Centro de Investigaciones y Proyectos Especiales, CIPE, Facultad de Finanzas, Gobierno y Relaciones Internacionales, Universidad Externado de Colombia, pp. 63-84.