Banco de México
Documentos de Investigación

Banco de México
Working Papers

N° 2006-13

**Estimación de la Volatilidad Integrada Utilizando
Retornos Absolutos en Alta Frecuencia**

Carla Ysusi
Banco de México

Diciembre 2006

La serie de Documentos de Investigación del Banco de México divulga resultados preliminares de trabajos de investigación económica realizados en el Banco de México con la finalidad de propiciar el intercambio y debate de ideas. El contenido de los Documentos de Investigación, así como las conclusiones que de ellos se derivan, son responsabilidad exclusiva de los autores y no reflejan necesariamente las del Banco de México.

The Working Papers series of Banco de México disseminates preliminary results of economic research conducted at Banco de México in order to promote the exchange and debate of ideas. The views and conclusions presented in the Working Papers are exclusively the responsibility of the authors and do not necessarily reflect those of Banco de México.

Estimación de la Volatilidad Integrada Utilizando Retornos Absolutos en Alta Frecuencia*

Carla Ysusi[†]
Banco de México

Resumen

Bajo modelos de volatilidad estocástica, la volatilidad spot integrada puede ser estimada con la variación absoluta realizada utilizando datos en alta frecuencia. Dadas las distribuciones asintóticas, la precisión de estos estimadores puede mejorarse a través de filtrado y suavizamiento. A pesar de que el uso de valores absolutos es empíricamente atractivo dado que son menos sensibles a posibles valores extremos, la variación absoluta realizada no es un estimador de la varianza integrada. Diferentes problemas pueden presentarse al usar un número finito de observaciones intra-día como se explica en este documento.

Palabras Clave: Variación cuadrática, Variación absoluta, Modelos de volatilidad estocástica, Semimartingala, Datos en alta frecuencia.

Abstract

When high-frequency data is available, in the context of a stochastic volatility model, realised absolute variation can estimate integrated spot volatility. A central limit theory enables us to do filtering and smoothing using model-based and model-free approaches in order to improve the precision of these estimators. Although the absolute values are empirically attractive as they are less sensitive to possible large movements in high-frequency data, realised absolute variation does not estimate integrated variance. Some problems arise when using a finite number of intra-day observations, as explained here.

Keywords: Quadratic variation, Absolute variation, Stochastic volatility models, Semimartingale, High-frequency data.

JEL Classification: C13, C51, G19.

*El autor agradece a Neil Shephard por sus consejos y apoyo, y a Valentina Corradi y Matthias Winkel por sus valiosas sugerencias.

[†] Dirección General de Investigación Económica. Email: cysusi@banxico.org.mx.

1 Introducción

Recientemente, los precios se han interpretado como realizaciones de los procesos de difusión en tiempo continuo. Registros completos de los precios para muchos activos financieros están disponibles en alta frecuencia, por lo tanto los modelos en tiempo continuo, se pueden calibrar. Más precisamente, en un proceso continuo semimartingala, la suma de los retornos cuadráticos en alta frecuencia estima la variación cuadrática. Es por ello que se puede utilizar la varianza realizada (la suma de retornos cuadráticos finitos intra-día) como estimador de la varianza integrada en un modelo de volatilidad estocástica. Este resultado ha sido estudiado ampliamente en la literatura reciente, por ejemplo, Andersen y Bollerslev (1998a), Barndorff-Nielsen y Shephard (2002) y Comte y Renault (1998). Sin embargo, esto es un resultado asintótico y los retornos infinitesimales no ocurren en la vida real. Se desarrolló una teoría asintótica del error de estimación para distinguir entre la verdadera variabilidad subyacente y el ruido de medición.

La varianza realizada ha sido utilizada en la econometría financiera por muchos años, por ejemplo, Rosenberg (1972), Merton (1980), Poterba y Summers (1986), Schwert (1989), Hsieh (1991), Zhou (1996), Taylor y Xu (1997), Christensen y Prabhala (1998), Andersen, Bollerslev, Diebold y Ebens (2001), Andersen, Bollerslev, Diebold y Labys (2001). La literatura reciente, utilizando la variación cuadrática para semimartingalas ha sido el desarrollo independiente y concurrente de Andersen y Bollerslev (1998a), Comte y Renault (1998) y Barndorff-Nielsen y Shephard (2001). Existen muchos otros documentos sobre la varianza realizada, los cuales se discuten en Dacorogna et al. (2001) y en las revisiones de Andersen, Bollerslev y Diebold (2005) y Barndorff-Nielsen y Shephard (2007).

Un enfoque alternativo para la varianza realizada sería usando la suma del valor absoluto de los incrementos en los log-precios intra-día, llamada la variación absoluta realizada, para estimar la volatilidad spot integrada de un modelo de volatilidad estocástica. Esto es empíricamente atractivo porque el uso de valores absolutos es menos sensible a posibles movimientos grandes en los datos de alta frecuencia. Taylor (1986) y Ding, Granger y Engle (1993) reconocieron que empíricamente los retornos absolutos son más persistentes que los retornos cuadráticos. Andersen y Bollerslev (1997) y Andersen y Bollerslev (1998b) estudiaron empíricamente las propiedades de la variación absoluta realizada de activos especulativos, sin embargo, abandonaron este enfoque en su trabajo subsecuente debido a la falta de una teoría adecuada. Ghysels, Santa-Clara y Valkanov (2003) y Forsberg y Ghysels (2006) retoman el interés en los retornos absolutos y proporcionan explicaciones empíricas y teóricas para el mejor desempeño de la variación absoluta realizada.

A pesar de sus ventajas empíricas, los retornos absolutos (la variación absoluta realizada) no se han estudiado a fondo. Por lo tanto, en el presente documento, dado un modelo de volatilidad estocástica donde los log-precios son una semimartingala continua, se utiliza la variación absoluta realizada para estimar la volatilidad spot integrada. De acuerdo con Barndorff-Nielsen, Nielsen, Shephard y Ysusi (2004), la teoría del límite correspondiente permitirá filtrar y suavizar utilizando un enfoque basado en un modelo y otro enfoque sin modelo, para mejorar la precisión de los estimadores.

Cuando se asume un proceso continuo subyacente para los log-precios, el uso de los datos en alta frecuencia para medir la volatilidad puede generar resultados inadecuados porque las observaciones discretas están contaminadas por los efectos de microestructura. Aunque la eficiencia de la variación absoluta realizada se mide con los resultados asintóticos, utilizar datos en la frecuencia más alta disponible no necesariamente será el mejor enfoque. Sin embargo, en el presente documento, se asumirá que las observaciones no están afectadas por el ruido de microestructura para

llegar a algunas conclusiones; se requieren más estudios para evaluar el efecto de este ruido en las estimaciones.

2 Marco y propiedades

Un modelo estándar en la economía financiera es el modelo de volatilidad estocástica (SV, por sus siglas en inglés) para los precios-log Y_t que sigue la ecuación

$$Y_t = \int_0^t a_u du + \int_0^t \sigma_s dW_s, t \geq 0, \quad (1)$$

donde $A_t = \int_0^t a_u du$. Se asume que los procesos σ_t y A_t son estocásticamente independientes del movimiento Browniano estándar W_t . Aquí σ_t es llamada la volatilidad instantánea o spot, σ_t^2 la varianza spot correspondiente y A_t el proceso de la media. Un ejemplo simple de este proceso es

$$A_t = \mu t + \beta \sigma_t^{2*}, \quad \text{donde} \quad \sigma_t^{2*} = \int_0^t \sigma_s^2 ds$$

El proceso σ_t^{2*} se llama la varianza integrada.

Más generalmente, se asume que A_t tiene una variación continua localmente acotada, y se define que $M_t = \int_0^t \sigma_s dW_s$, con la condición adicional de que $\int_0^t \sigma_s^2 ds < \infty$ para todos t . Lo anterior garantiza que M_t es una martingala local. Por lo tanto, la ecuación original (1) se puede descomponer como

$$Y_t = A_t + M_t$$

Bajo estos supuestos, Y_t es una semimartingala continua (véase Protter (1990)).

Es esencial definir una versión discretizada de Y_t con base en los intervalos de tiempo de longitud $\delta > 0$. Dado el marco previo, los retornos δ se denotan como:

$$y_j = Y_{j\delta} - Y_{(j-1)\delta} \quad j = 1, 2, 3, \dots, \lfloor t/\delta \rfloor$$

Así, independientemente del modelo para la volatilidad, si A_t y σ_t son estocásticamente independientes de W_t , esto implica que

$$y_j \mid a_j, \nu_j^2 \sim N(a_j, \nu_j^2) \quad (2)$$

donde $a_j = A_{j\delta} - A_{(j-1)\delta}$ y $\nu_j^2 = \sigma_{j\delta}^{2*} - \sigma_{(j-1)\delta}^{2*}$. Normalmente, a_j se llama la verdadera media y ν_j^2 es la verdadera varianza.

Uno de los aspectos más importantes de las semimartingalas es la variación cuadrática (QV, por sus siglas en inglés), definida como

$$[Y]_t = p \lim_{n \rightarrow \infty} \sum_{j=1}^n (Y_{t_j} - Y_{t_{j-1}})^2,$$

para cualquier secuencia de particiones $t_0^{(n)} = 0 < t_1^{(n)} < \dots < t_n^{(n)} = t$ con $\sup_j \{t_j^{(n)} - t_{j-1}^{(n)}\} \rightarrow 0$ cuando $n \rightarrow \infty$.

Ya que se asume que A_t es continuo y con variación finita, se obtiene que

$$[Y]_t = [A]_t + 2[A, M]_t + [M]_t = [M]_t = \int_0^t \sigma_u^2 du$$

donde

$$[X, Y]_t = p \lim_{n \rightarrow \infty} \sum_{j=1}^n (X_{t_j} - X_{t_{j-1}}) (Y_{t_j} - Y_{t_{j-1}}) .$$

Esta ecuación se mantiene, dado que la variación cuadrática de cualquier proceso continuo con variación localmente acotada es cero (véase Protter (1990)).

La generalización de la variación cuadrática es

$$[Y]_t^{[r]} = p \lim_{n \rightarrow \infty} \sum_{j=1}^n |Y_{t_j} - Y_{t_{j-1}}|^r ,$$

para cualquier secuencia de particiones $t_0^{(n)} = 0 < t_1^{(n)} < \dots < t_n^{(n)} = t$ con $\sup_j \{t_j^{(n)} - t_{j-1}^{(n)}\} \rightarrow 0$ cuando $n \rightarrow \infty$.

Ahora, basado en la volatilidad spot σ_t , la volatilidad spot integrada se puede definir alternativamente como

$$\sigma_t^{[1]*} = \int_0^t \sigma_s ds$$

y la verdadera volatilidad como

$$\nu_j^{[1]} = \sigma_{j\delta}^{[1]*} - \sigma_{(j-1)\delta}^{[1]*} \quad j = 1, 2, 3, \dots, \lfloor t/\delta \rfloor .$$

Es importante notar que la verdadera volatilidad es diferente a la raíz cuadrada de la verdadera varianza, así que $(\nu_j^{[1]})^2 \neq \nu_j^2$.

Más información sobre los modelos SV se encuentra en, por ejemplo, Shiryaev (1999) o Shephard (2005).

2.1 Variación absoluta realizada

El proceso de la variación absoluta realizada se define como

$$[Y_\delta]_t^{[1]} = \sum_{j=1}^{\lfloor t/\delta \rfloor} |y|_j$$

Barndorff-Nielsen y Shephard (2003) derivaron un teorema de límite para el proceso de variación realizada elevada a la potencia

$$[Y_\delta]_t^{[r]} = \sum_{j=1}^{\lfloor t/\delta \rfloor} |y|_j^r$$

cuando $\delta \downarrow 0$. Dado que la variación realizada absoluta es un caso especial de la variación realizada elevada a la potencia, $r = 1$, se puede utilizar el resultado asintótico para estudiar las propiedades del error de estimación de la volatilidad spot integrada.

Barndorff-Nielsen y Shephard (2003) presentan el Teorema Central del Límite (TCL, ó *CLT* sus siglas en inglés) para la variación realizada elevada a la potencia bajo condiciones relativamente restrictivas. Barndorff-Nielsen, Graversen, Jacod, Podolskij y Shephard (2005) y Barndorff-Nielsen, Graversen, Jacod y Shephard (2006) presentan algunos resultados generales del TCL para la variación realizada elevada a la potencia y bipoder, los cuales se comprobaron bajo supuestos mucho menos restrictivos.

Para el modelo SV (1), donde A_t es de variación localmente acotada, $\int_0^t \sigma_u^2 du < \infty$ y σ_t es *càdlàg*, resulta que para $\delta \downarrow 0$

$$\mu_r^{-1} \delta^{1-r/2} [Y_\delta]_t^{[r]} \xrightarrow{p} \int_0^t \sigma_s^r ds$$

y

$$\frac{\mu_r^{-1} \delta^{1-r/2} [Y_\delta]_t^{[r]} - \int_0^t \sigma_s^r ds}{\mu_r^{-1} \delta^{1-r/2} \sqrt{\mu_{2r}^{-1} v_r [Y_\delta]_t^{[2r]}}} \xrightarrow{L} N(0, 1)$$

donde $\mu_r = E(|X|^r)$ y $v_r = Var(|X|^r)$ con $X \sim N(0, 1)$.

En particular, esto implica que

$$\frac{d[Y]_t^{[1]}}{dt} = \sigma_t.$$

Con estas ecuaciones, se puede obtener el resultado relevante para el caso ($r = 1$) donde

$$\frac{\sqrt{\delta}}{\mu_1} [Y_\delta]_t^{[1]} \xrightarrow{p} [Y]_t^{[1]}$$

donde $\mu_1 = \sqrt{2/\pi}$ y

$$\frac{\frac{\sqrt{\delta}}{\mu_1} [Y_\delta]_t^{[1]} - [Y]_t^{[1]}}{\sqrt{\delta(\mu_1^{-2} - 1) [Y_\delta]_t^{[2]}}} \xrightarrow{L} N(0, 1)$$

Para las series diarias, se supone que existen $\lfloor t/\delta \rfloor = M$ observaciones intra- h , durante cada periodo de tiempo con h fijo (h denota el periodo de un día), definidas como

$$y_{j,i} = Y_{(i-1)h+j\delta} - Y_{(i-1)h+(j-1)\delta}$$

para el j -ésimo retorno intra- h para el i -ésimo periodo. Entonces, la variación absoluta realizada, es decir, una suma escalada del valor absoluto de los cambios intra-día de los log-precios, se definirá aquí como

$$[Y_M]_i^{[1]} = \frac{1}{\sqrt{M\mu_1}} \sum_{j=1}^M |y_{j,i}|$$

y por lo tanto, $[Y_M]_i^{[1]} \rightarrow \sigma_{ih}^{*[1]} - \sigma_{(i-1)h}^{*[1]}$ cuando $M \rightarrow \infty$. Entonces, $\widehat{\nu_i^{[1]}} = [Y_M]_i^{[1]}$ converge a $\nu_i^{[1]}$.

Consecuentemente, se puede utilizar la variación absoluta realizada como estimador de la volatilidad real.

Utilizando los precios INTEL intra-día, se traza la variación absoluta realizada, usando los retornos de cinco y sesenta minutos, y las funciones de autocorrelación (FAC, ó *ACF* sus siglas en inglés) correspondientes (Gráfica 1). Cuando M disminuye, las series se vuelen más erráticas. También, la autocorrelación es mayor cuando el valor de M es alto, y muestra una caída más lenta.

Gráfica 1. Variación absoluta realizada, calculada a partir de los retornos de 5 y 60 minutos de INTEL, y de sus funciones de autocorrelación.

2.2 Varianza realizada

El proceso de la variación cuadrática realizada se define como

$$[Y_\delta]_t^{[2]} = \sum_{j=1}^{\lfloor t/\delta \rfloor} y_{j,\delta}^2, \quad \text{entonces mientras } \delta \downarrow 0 \quad [Y_\delta]_t^{[2]} \xrightarrow{p} [Y]_t$$

Para las series diarias, la varianza realizada se define por la suma de los cuadrados de estos M retornos intra-periodo y, bajo semimartingala, cuando $M \rightarrow \infty$, entonces

$$[Y_M]_i^{[2]} = \sum_{j=1}^M y_{j,i}^2 \xrightarrow{p} v_i^2 = [Y]_{ih} - [Y]_{(i-1)h}$$

por lo tanto, se puede utilizar la varianza realizada como estimador de la verdadera varianza.

En Barndorff-Nielsen y Shephard (2002), Barndorff-Nielsen y Shephard (2005) y Barndorff-Nielsen y Shephard (2004b), la teoría previa se ha extendido a un Teorema Central del Límite. En estos documentos, el TCL se presenta bajo supuestos relativamente restrictivos. Recientemente, Barndorff-Nielsen, Graversen, Jacod, Podolskij y Shephard (2005) y Barndorff-Nielsen, Graversen,

Jacod y Shephard (2006) presentan condiciones más débiles para el proceso de log-precios, asegurando que se mantiene el TCL. Para el modelo SV (1), cuando $\delta \downarrow 0$

$$\frac{\delta^{-1/2}([Y_\delta]_t^{[2]} - [Y]_t)}{\sqrt{2 \int_0^t \sigma_s^4 ds}} \xrightarrow{L} N(0, 1)$$

bajo los supuestos de que A_t es de variación localmente acotada, $\int_0^t \sigma_u^2 du < \infty$ y que σ_t es càdlàg.

Gráfica 2. Series de la variación realizada, utilizando los retornos a 5 y 60 minutos de INTEL y sus funciones de autocorrelación.

La Gráfica 2 ilustra la serie de tiempo de la varianza realizada para el conjunto de datos INTEL. En la parte superior de esta gráfica, las varianzas realizadas y sus ACFs se computaron utilizando M=78, que corresponde a utilizar datos de retornos de cinco minutos basados en el conjunto de datos INTEL. Las varianzas realizadas computadas con M=6, retornos de sesenta minutos, y las ACF correspondientes se presentan en la parte inferior de la gráfica. Se puede observar que la serie de varianza realizada, calculada utilizando retornos de cinco minutos (M=78) es mucho menos errática que la serie calculada con los retornos de sesenta minutos (M=6). Los correlogramas tienen el comportamiento usual de una caída lenta y empiezan en un nivel bastante bajo para M=6. Al comparar las ACF de la variación absoluta realizada con las de la varianza realizada, se puede notar que la ACF de la variación absoluta realizada es marginalmente más fuerte.

3 Estimaciones con la Variación Absoluta Realizada

Los retornos absolutos tienen propiedades empíricas deseables, sin embargo, se requiere más investigación sobre su uso como estimadores de la volatilidad spot integrada. En esta sección, se definirá un modelo y se estimará la volatilidad spot integrada utilizando la variación absoluta realizada. Barndorff-Nielsen, Nielsen, Shephard y Ysusi (2004) realizaron un análisis similar para la varianza realizada, pero también se tienen que estudiar los retornos absolutos.

Cuando se estima $\nu_i^{[1]}$ con la variación absoluta realizada, la varianza del error es bastante alta, incluso cuando se usan valores altos de M , por lo tanto, se debería establecer un modelo. Un modelo lineal para $\nu_j^{[1]}$ se puede definir como

$$\nu_{s:p}^{[1]} = cE\left(v_{s:p}^{[1]}\right) + A[Y_M]_{s:p}^{[1]}$$

donde

$$[Y_M]_{s:p}^{[1]} = \frac{1}{\sqrt{M}\mu_1} \left(\sum_{j=1}^M |y_{j,s}|, \sum_{j=1}^M |y_{j,s+1}|, \dots, \sum_{j=1}^M |y_{j,p}| \right)'$$

y

$$\nu_{s:p}^{[1]} = \left(\nu_s^{[1]}, \nu_{s+1}^{[1]}, \dots, \nu_p^{[1]} \right)'$$

Sabemos que

$$\sqrt{\frac{M}{h}} \left([Y_M]_{s:p}^{[1]} - v_{s:p}^{[1]} \right) |v_{s:p}^2 \xrightarrow{L} N\left(0, \left(\frac{\pi}{2} - 1\right) A \text{diag}(v_{s:p}^2) A'\right)$$

Por lo tanto, para la variación absoluta realizada se puede obtener que

$$\sqrt{\frac{M}{h}} \left(A[Y_M]_{s:p}^{[1]} - Av_{s:p}^{[1]} \right) |v_{s:p}^2 \xrightarrow{L} N\left(0, \left(\frac{\pi}{2} - 1\right) A \text{diag}(v_{s:p}^2) A'\right)$$

Asumiendo que las variaciones absolutas realizadas son un proceso estacionario en covarianza, el estimador ponderado de mínimos cuadrados da en este caso que

$$\begin{aligned} \hat{c} &= (I - \hat{A})\iota \\ \hat{A} &= Cov(v_{s:p}^{[1]}) \left(Cov(v_{s:p}^{[1]}) + \left(\frac{\pi}{2} - 1\right) \frac{E(v_i^2)h}{M} I \right)^{-1} \end{aligned}$$

Los estimadores se pueden obtener sin modelo o basados en un modelo. Primero, se observa el enfoque sin modelo.

3.1 Método basado en la muestra

Cuando se utilizan los promedios empíricos, si se tiene un proceso estacionario en covarianza de las variaciones absolutas realizadas y un proceso ergódico diario, entonces

$$\frac{1}{T} \sum_{i=1}^T [Y_M]_i^{[1]} \xrightarrow{p} E(v_i^{[1]})$$

y

$$\left(\frac{1}{T} \sum_{i=1}^T \sum_{j=1}^M y_{j,i}^2 \right) \xrightarrow{p} E(v_i^2)$$

cuando M y T tienden infinito. Con estos resultados, las estimaciones se pueden obtener como

$$\hat{A} = \left(Cov([Y_M]_{s:p}^{[1]}) - \left(\frac{\pi}{2} - 1 \right) \frac{E(v_i^2)h}{M} I \right) \left(Cov([Y_M]_{s:p}^{[1]}) \right)^{-1}$$

$Cov([Y_M]_{s:p}^{[1]})$ se puede calcular fácilmente con los datos, y dado que se trabaja con datos diarios, se fija $h = 1$.

Los primeros resultados se obtienen utilizando solamente un valor de la variación absoluta realizada en el modelo, $s = p = i$. En el Cuadro 1 se puede apreciar que, como se espera, cuando M aumenta, la ponderación dada a la variación absoluta realizada se vuelve más importante. La diferencia en las ponderaciones, cuando se usan los retornos de cinco minutos en comparación con el uso de los retornos de sesenta minutos, es muy ilustrativa en este sentido.

M	min	$\hat{c}$	$\hat{A}$
78	5	0.0802	0.9197
39	10	0.1413	0.8586
26	15	0.2050	0.7949
19	20	0.2513	0.7486
13	30	0.3652	0.6347
6	60	0.5536	0.4463

Cuadro 1. Pesos estimados para el cálculo de $\nu_i^{[1]}$.

3.1.1 Modelos con rezagos y adelantados

En esta sección, con un enfoque dinámico, iniciando con el modelo más sencillo, se estima $\nu_{s:p}^{[1]}$ utilizando un rezago, un componente adelantado y la variación absoluta realizada contemporánea.

Cuando se trabaja con los retornos de cinco minutos, se obtienen las siguientes matrices.

$$\hat{A} = \begin{pmatrix} 0.86617 & 0.07172 & 0.01856 \\ 0.07172 & 0.83031 & 0.07172 \\ 0.01856 & 0.07172 & 0.86617 \end{pmatrix} \quad \hat{c} = \begin{pmatrix} 0.04354 \\ 0.02624 \\ 0.04354 \end{pmatrix}$$

entonces, las ponderaciones se obtienen del segundo renglón de las matrices, dado el siguiente estimador

$$\widehat{\nu_i^{[1]}} = 0.026E(v_i^{[1]}) + 0.071 [Y_{78}]_{i-1}^{[1]} + 0.830 [Y_{78}]_i^{[1]} + 0.071 [Y_{78}]_{i+1}^{[1]}$$

Se puede observar que el coeficiente de la variación absoluta realizada contemporánea es mucho mayor que los del rezago y del adelanto. Sin embargo, si se compara este estimador con el previo sin rezago y sin adelanto, la ponderación de la expectativa es menor y las ponderaciones dadas en total a la variación absoluta realizada (rezago, componente contemporáneo y adelantado) han aumentado.

Cuando se calcula la variación absoluta realizada con los retornos de sesenta minutos, el estimador es

$$\widehat{\nu_i^{[1]}} = 0.331E(v_i^{[1]}) + 0.159 [Y_6]_{i-1}^{[1]} + 0.349 [Y_6]_i^{[1]} + 0.159 [Y_6]_{i+1}^{[1]}$$

Aún, otra vez, se puede observar que, mientras menor sea el valor de M , menor será la ponderación dada a la variación absoluta realizada. Al utilizar un rezago y un componente adelantado, la ponderación en la expectativa de $\nu_i^{[1]}$ disminuye, y la ponderación dada en total a la variación absoluta realizada aumenta en comparación con la estimación sin rezago y sin adelanto, independientemente del valor de M .

Se puede aumentar el número de componentes rezagados y componentes adelantados en el modelo. Cuando se utilizan dos rezagos y dos adelantos para la estimación, se obtiene una matriz 5x5 para $\hat{A}$ y un vector 5x1 para $\hat{c}$. Utilizando su tercer renglón, se obtienen los siguientes modelos cuando se usan los retornos de cinco y sesenta minutos, consecutivamente

$$\begin{aligned}\widehat{\nu_i^{[1]}} &= 0.015E(v_i^{[1]}) + 0.016 [Y_{78}]_{i-2}^{[1]} + 0.062 [Y_{78}]_{i-1}^{[1]} + 0.826 [Y_{78}]_i^{[1]} + 0.062 [Y_{78}]_{i+1}^{[1]} + 0.016 [Y_{78}]_{i+2}^{[1]} \\ \widehat{\nu_i^{[1]}} &= 0.253E(v_i^{[1]}) + 0.073 [Y_6]_{i-2}^{[1]} + 0.133 [Y_6]_{i-1}^{[1]} + 0.332 [Y_6]_i^{[1]} + 0.133 [Y_6]_{i+1}^{[1]} + 0.073 [Y_6]_{i+2}^{[1]}\end{aligned}$$

Aquí, se nota nuevamente que, mientras mayor sea el valor de M , mayor será la ponderación dada a la variación absoluta realizada contemporánea. También, mientras más cercanos estén los rezagos y los adelantos a la variación absoluta realizada contemporánea, mayor será la ponderación dada a ellos. Las ponderaciones de la expectativa incondicional son incluso menores porque la verdadera volatilidad se puede explicar mejor con los rezagos.

En la Gráfica 3, las ponderaciones del modelo, utilizando tres rezagos, tres componentes adelantados y la variación absoluta realizada contemporánea, se muestran para distintos valores de M . Se nota qué tan rápido las ponderaciones se enfocan en la observación contemporánea de la variación absoluta realizada cuando M aumenta.

Gráfica 3. Pesos para estimar la volatilidad real usando 3 rezagos y 3 adelantos para diferentes valores de M .

3.1.2 Logaritmos

Se puede establecer un análisis similar utilizando las log-variaciones absolutas realizadas. Los resultados asintóticos relevantes, obtenidos con el método Delta, son los siguientes

$$\sqrt{\frac{M}{h}} \left(A \log [Y_M]_{s:p}^{[1]} - A \log v_{s:p}^{[1]} \right) | v_{s:p}^2, v_{s:p}^{[1]} \xrightarrow{L} N(0, (\frac{\pi}{2} - 1) A E \begin{pmatrix} \frac{v_s^2}{(v_s^{[1]})^2} & 0 & \dots & 0 \\ 0 & \ddots & & 0 \\ \dots & & & \\ 0 & 0 & & \frac{v_p^2}{(v_p^{[1]})^2} \end{pmatrix} A')$$

entonces, se puede usar el hecho de que

$$\left(\frac{1}{T} \sum_{i=1}^T \frac{\sum_{j=1}^M y_{j,i}^2}{\left(\frac{1}{\sqrt{M\mu}} \sum_{j=1}^M |y_{j,i}|^2 \right)} \right) \xrightarrow{P} E \left(\frac{v_i^2}{(v_i^{[1]})^2} \right)$$

y que

$$\frac{1}{T} \sum_{i=1}^T \log [Y_M]_i^{[1]} \xrightarrow{P} E(\log v_i^{[1]})$$

para obtener los estimadores.

Esto permite construir la estimación

$$\widehat{\log \nu_{s:p}^{[1]}} = \widehat{c} E(\log v_{s:p}^{[1]}) + \widehat{A} \log [Y_M]_{s:p}^{[1]}$$

y dado que las variaciones absolutas realizadas son un proceso estacionario en covarianza, entonces, el conjunto del estimador ponderado de mínimos cuadrados

$$\widehat{c} = (I - \widehat{A})_\iota$$

y

$$\widehat{A} = Cov(\log \nu_{s:p}^{[1]}) \left(Cov(\log v_{s:p}^{[1]}) + (\frac{\pi}{2} - 1) \frac{h}{M} E \left(\frac{v_i^2}{(v_i^{[1]})^2} \right) I \right)^{-1}$$

Primero, al asignar $s = p = i$, es decir, utilizando solamente la variación absoluta realizada contemporánea, y utilizando los logaritmos, se obtienen los resultados mostrados en el Cuadro 2. Como antes, mientras menor sean los valores utilizados de M , menor será la ponderación dada a la variación absoluta realizada. En comparación con el Cuadro 1, cuando se usan los logaritmos y el valor de M es pequeño, se encuentran mayores valores para la ponderación de la log-variación absoluta realizada. Nótese que las ponderaciones aumentan cuando M tiene un valor pequeño.

M	min	$\widehat{c}$	$\widehat{A}$
78	5	0.0802	0.9137
39	10	0.1465	0.8534
26	15	0.2045	0.7954
19	20	0.2633	0.7366
13	30	0.3337	0.6662
6	60	0.4718	0.5281
Cuadro 2. Pesos estimados para el cálculo del logaritmo de $\nu_i^{[1]}$.			

Al utilizar un rezago y un componente adelantado en el log-modelo, se pueden obtener las mismas conclusiones.

Para los retornos de cinco minutos, se obtiene el modelo

$$\widehat{\nu_i^{[1]}} = 0.032E(\log v_i^{[1]}) + 0.067 \log [Y_{78}]_{i-1}^{[1]} + 0.833 \log [Y_{78}]_i^{[1]} + 0.067 \log [Y_{78}]_{i+1}^{[1]}$$

y para sesenta minutos

$$\widehat{\nu_i^{[1]}} = 0.316E(\log v_i^{[1]}) + 0.101 \log [Y_6]_{i-1}^{[1]} + 0.479 \log [Y_6]_i^{[1]} + 0.101 \log [Y_6]_{i+1}^{[1]}$$

En la Gráfica 4, se muestran las ponderaciones del log-modelo utilizando tres rezagos, tres adelantos y la observación contemporánea para distintos valores de M . Cuando se utilizan los logaritmos, las ponderaciones dadas para valores pequeños de M son mucho mayores que las obtenidas para el estimador basado en la variación absoluta realizada simple (sin logaritmo) (véase la Gráfica 3 para la comparación).

Gráfica 4. Pesos para estimar $\log(\nu_i^{[1]})$ usando 3 rezagos y 3 adelantos para diferentes valores de M .

3.2 Estimación basada en un modelo

Para construir un estimador de la volatilidad real basado en un modelo, se tienen que derivar los primeros y segundos momentos de $\nu_i^{[1]}$. Si ξ es la media de σ_t , w es su varianza, y r es su función de autocorrelación, entonces

$$E(\nu_i^{[1]}) = h\xi, \quad Var(\nu_i^{[1]}) = 2\omega^2 r_h^{**} \quad y \quad Cov\{v_i^{[1]}, v_{i+s}^{[1]}\} = \omega^2 \diamond r_{hs}^{**}$$

donde

$$\diamond r_{hs}^{**} = r_{s+h}^{**} - 2r_s^{**} + r_{s-h}^{**} \quad r_t^* = \int_0^t r_u du$$

y

$$r_t^{**} = \int_0^t r_u^* du$$

Entonces, con las propiedades de segundo orden de σ_t , se pueden establecer completamente las propiedades de segundo orden de $\nu_i^{[1]}$.

En un contexto de un modelo de volatilidad estocástica, la variación absoluta realizada se puede escribir como

$$[Y_M]_i^{[1]} = \nu_i^{[1]} + u_i,$$

así

$$u_i = [Y_M]_i^{[1]} - \nu_i^{[1]} = \frac{1}{\sqrt{M}\mu} \sum_{j=1}^M |\varepsilon_{j,i}| \sqrt{v_{j,i}^2} - \sum_{j=1}^M \nu_{j,i}^{[1]}$$

donde

$$\begin{aligned} \varepsilon_{j,i} &\stackrel{iid}{\sim} N(0, 1) \\ \nu_{j,i}^2 &= \sigma_{(i-1)h + \frac{jh}{M}}^{2*} - \sigma_{(i-1)h + \frac{(j-1)h}{M}}^{2*}, \\ \nu_{j,i}^{[1]} &= \sigma_{(i-1)h + \frac{jh}{M}}^{[1]*} - \sigma_{(i-1)h + \frac{(j-1)h}{M}}^{[1]*}, \\ \sigma_t^{[1]*} &= \int_0^t \sigma_u du \quad y \quad \sigma_t^{2*} = \int_0^t \sigma_u^2 du. \end{aligned}$$

Como se mencionó antes $\sqrt{\nu_{j,i}^2} \neq \nu_{j,i}^{[1]}$, entonces

$$E(\sqrt{M}u_i | v_i^{[1]}) \neq 0$$

Sin embargo, $\sqrt{M}u_i \xrightarrow{L} MN(0, (\frac{\pi}{2} - 1)h \int_{(i-1)h}^{ih} \sigma_u^2 du)$, por lo tanto,

$$E(\sqrt{M}u_i | v_i^{[1]}) = o(1).$$

Se pueden encontrar las propiedades de segundo orden de $[Y_M]_i^{[1]}$. Primero, sea

$$\begin{aligned} E([Y_M]_i^{[1]}) &= E(v_i^{[1]}) + E(u_i) = h\xi + \frac{1}{\sqrt{M}}o(1) \\ Var([Y_M]_i^{[1]}) &= Var(v_i^{[1]}) + Var(u_i) + o(1) \\ Cov([Y_M]_i^{[1]}, [Y_M]_{i+s}^{[1]}) &= Cov(v_i^{[1]}, v_{i+s}^{[1]}) + o(1) \end{aligned}$$

Falta $Var(u_i)$ para determinar estas propiedades de segundo orden. De igual manera, dado el hecho de que

$$\sqrt{M}u_i \xrightarrow{L} MN\left(0, \left(\frac{\pi}{2} - 1\right)h \int_{(i-1)h}^{ih} \sigma_u^2 du\right)$$

se obtiene que

$$\begin{aligned}
Var(\sqrt{M}u_i) &= (\frac{\pi}{2} - 1)hE\left(\int_{(i-1)h}^{ih} \sigma_u^2 du\right) + o(1) \\
&= (\frac{\pi}{2} - 1)h \int_{(i-1)h}^{ih} E(\sigma_u^2) du + o(1) \\
&= (\frac{\pi}{2} - 1)h^2 E(\sigma_u^2) + o(1) \\
&= (\frac{\pi}{2} - 1)h^2 (Var(\sigma_u) + E^2(\sigma_u)) + o(1)
\end{aligned}$$

entonces,

$$Var(\sqrt{M}u_i) = (\frac{\pi}{2} - 1)h^2(\omega^2 + \xi^2) + o(1)$$

Con estos resultados, la varianza de la variación realizada y su covarianza pueden ser establecidas,

$$\begin{aligned}
Var([Y_M]_i^{[1]}) &= (\frac{\pi}{2} - 1)h^2(\omega^2 + \xi^2) + 2\omega^2 r_h^{**} + \frac{1}{M}o(1) \\
Cov([Y_M]_i^{[1]}, [Y_M]_{i+s}^{[1]}) &= \omega^2 \diamond r_{sh}^{**} + o(1)
\end{aligned}$$

3.2.1 Modelo

Similar al caso de varianza realizada, para modelar la volatilidad estocástica, se usará un proceso con una función de autocorrelación como $r_t = \exp(-\lambda |t|)$. El proceso, que es la solución para la ecuación estocástica diferencial

$$d\sigma_t = -\lambda\sigma_t dt + dz_{\lambda t} \quad (3)$$

donde z_t es un proceso Levy con incrementos no negativos, tiene la ACF previa.

Para este proceso $r_t^{**} = -\lambda^{-2}\{e^{-\lambda t} - 1 + \lambda t\}$ y $\diamond r_{sh}^{**} = \lambda^{-2}(1 - e^{-\lambda t})^2 e^{-\lambda(s-1)h}$, $s > 0$; implicando que los momentos asintóticos son

$$\begin{aligned}
E(\nu_i^{[1]}) &= \xi h \\
Var(\nu_i^{[1]}) &= \frac{2\omega^2}{\lambda^2} (e^{-\lambda h} - 1 + \lambda h) \\
Cor(\nu_i^{[1]}, \nu_{i+s}^{[1]}) &= \frac{(1 - e^{-\lambda h})^2 (e^{-\lambda(s-1)h})}{2(e^{-\lambda h} - 1 + \lambda h)}
\end{aligned}$$

El modelo de autocorrelación para $\nu_i^{[1]}$ es aquel para un modelo ARMA(1,1), entonces se puede usar la siguiente representación lineal de estado-espacio:

$$\begin{aligned}
[Y_M]_i^{[1]} &= \xi h + (1 \ 0)\alpha_i + \sigma_u v_{1i} \\
\alpha_{i+1} &= \begin{pmatrix} \phi & 1 \\ 0 & 0 \end{pmatrix} \alpha_i + \begin{pmatrix} \sigma_\sigma \\ \sigma_\sigma \theta \end{pmatrix} v_{2i}
\end{aligned}$$

donde

$$\alpha_i = \nu_i^{[1]} - \xi h \quad y \quad u_i = \sigma_u v_{1i}$$

En este caso, v_i es una secuencia con media cero y ruido blanco con una matriz identidad de covarianza; ϕ , θ y σ_σ son la raíz autorregresiva, la raíz de media móvil y la varianza de un choque en el proceso. Finalmente σ_u^2 se obtiene del resultado asintótico para la $Var(u_i)$. Se puede utilizar el filtro Kalman para predecir y suavizar $\nu_i^{[1]}$.

Nuevamente, solamente uno de los modelos previos de volatilidad es demasiado simple para ajustar la dependencia larga que tienen las series de tiempo financieras, por lo tanto, se superpondrán los procesos para describir σ_u . En este caso, se obtiene

$$\sigma_t = \sum_{j=1}^J \sigma_t^{(j)} \quad y \quad \sum_{j=1}^J w_j = 1$$

donde w_j son las ponderaciones que tienen que ser mayores o iguales a cero y $\sigma_t^{(j)}$ son los procesos con memoria λ_j .

Utilizando el modelo previo, se puede estimar $\nu_i^{[1]}$ para el conjunto de datos INTEL. El modelo se ajustó utilizando los retornos de cinco, quince y sesenta minutos, determinando el número de procesos necesarios en la superposición y estimando los parámetros.

Cuando se utilizan los retornos de cinco minutos, la AFC empírica se describe mejor con la superposición de tres procesos (Gráfica 5); agregando un proceso adicional para obtener una superposición de cuatro procesos no contribuye. No obstante, dos procesos no fueron suficientes para describir la AFC, entonces, este tercer componente es esencial. El Cuadro 3 muestra cómo dos de los procesos tienen mucha memoria (valores bajos de λ) y el otro proceso tiene una memoria muy baja (valor alto de λ). Existe una diferencia importante en el estadístico Box-Pierce de los modelos basados en uno o dos procesos. Se puede notar que el estimador basado en un modelo con solamente un proceso no presenta un buen ajuste. No es hasta la superposición de tres procesos, cuando se captura la dependencia larga de los datos.

El modelo también se ajusta para valores más pequeños de M , primero para los retornos de quince minutos. También en este caso, la superposición de tres procesos es el modelo que se ajusta mejor. Dos de los componentes tienen valores bajos de λ y solamente un componente tiene un valor alto, explicando la dependencia larga. Lo mismo ocurre cuando se utilizan los retornos de sesenta minutos; el mejor ajuste se encuentra cuando se utilizan los tres procesos. También, dos de los componentes tienen una varianza persistente (valor bajo de λ), y el ajuste no se mejora cuando se agrega un cuarto proceso, pero sí se mejora cuando cambia de dos a tres procesos.

Gráfica 5. Usando $M = 78$, la FAC de $[Y_M]_i^{[1]}$ y la versión ajustada para varios valores de J .

	ξ	ω^2	λ_1	λ_2	λ_3	λ_4	w_1	w_2	w_3	L_Q	BP_{20}
J = 1	3.41	5.39	0.64	-	-	-	1.00	-	-	-11,191	53.61
J = 2	3.48	5.08	0.03	0.57	-	-	0.18	-	-	-10,834	13.41
J = 3	3.48	7.02	0.04	0.61	291.9	-	0.15	0.26	-	-10,804	10.88
J = 4	3.49	7.05	0.04	0.61	57.3	291.9	0.19	0.32	0.42	-10,804	10.89

Cuadro 3. Ajuste de la superposición del proceso de volatilidad J para un modelo VE basado en absoluta realizada usando M=78.
Parámetros, Cuasi-Verosimilitud (L_Q) y estadístico de Box-Pierce (BP).

3.2.2 Comparación

Se puede comparar la serie suavizada del enfoque sin modelo y del enfoque basado en un modelo. El Cuadro 4 muestra la correlación entre los estimadores basados en el modelo y los estimadores sin modelo con distintos números de rezagos y componentes adelantados y para distintos valores de M . Cuando el número de rezagos y adelantos aumenta, los estimadores se vuelven más correlacionados, y también cuando M aumenta, la conexión entre los estimadores es más fuerte. En la Gráfica 6 se muestran los primeros 200 valores estimados de $\nu_i^{[1]}$ de las series de tiempo. Las series de tiempo corresponden a los estimadores sin modelo, utilizando dos rezagos y dos componentes adelantados y el estimador suavizado basado en el modelo. Se puede apreciar que existe una conexión fuerte entre los dos estimadores.

	5 min.	15 min.	60 min.
s = p	0.9587	0.8941	0.7838
1 rezago, 1 adelanto	0.9805	0.9612	0.9508
2 rezagos, 2 adelantos	0.9811	0.9692	0.9772
logaritmos	0.9586	0.8925	0.7702
logaritmos 1 rezago, 1 adelanto	0.9781	0.9507	0.8836

Cuadro 4. Correlación entre los estimadores sin modelos y los basados en el modelo.

Gráfica 6. Estimadores de $\nu_i^{[1]}$, sin modelo usando 2 rezagos y 2 adelantos, y en base al modelo, para M=78 y M=6.

3.3 Retornos absolutos vs. retornos cuadráticos

Es bien conocido que los retornos absolutos pueden generar mejores resultados que los retornos cuadráticos para la estimación de la varianza integrada. Al aplicar el modelo previo equivalente, el filtro y el suavizamiento previo equivalente, pero para la varianza realizada (como en Barndorff-Nielsen, Nielsen, Shephard y Ysusi (2004)) utilizando la base de datos INTEL, la variación absoluta realizada parece ser más robusta que la variación realizada. Esto es consistente con Forsberg y Ghysels (2006) que proporcionan una explicación teórica para el hecho de que la variación absoluta realizada supere la varianza realizada cuando se estima la volatilidad spot integrada y la varianza integrada respectivamente.

Utilizando momentos poblacionales, estos autores prueban que el uso de los valores absolutos genera una mayor persistencia. Cuando se utilizan datos en alta frecuencia para la estimación, el sesgo debido a errores de medición es mucho menor con retornos absolutos (variación absoluta realizada) que con retornos cuadráticos (varianza realizada). Los autores estudiaron el efecto de brinco discretos en el proceso de log-precios y concluyeron que los retornos absolutos fueron inmunes a su presencia. Destacaron tres razones por las que los retornos absolutos más persistentes que los retornos cuadráticos: 1) las características deseables de predictibilidad poblacional, 2) mejor comportamiento de muestra del error y 3) inmunes a los brinco.

De esto, se puede concluir que la variación absoluta realizada es un mejor estimador de la volatilidad spot integrada que la varianza realizada para la varianza integrada. Sin embargo, el objetivo principal es la estimación de la varianza integrada. Se comparan las distribuciones asintóticas del error para el proceso de varianza realizada y el proceso de variación absoluta realizada cuadrada, así como sus transformaciones logarítmicas. En el primer caso, se compara

$$[Y_\delta]_t^{[2]} - \int_0^t \sigma_u^2 du \xrightarrow{L} MN(0, 2/\delta \int_0^t \sigma_u^4 du)$$

y

$$\left([Y_\delta]_t^{[1]}\right)^2 - \left(\int_0^t \sigma_u du\right)^2 \xrightarrow{L} MN\left(0, \frac{4}{\delta}\left(\frac{\pi}{2} - 1\right) \int_0^t \sigma_u^2 du \left(\int_0^t \sigma_u du\right)^2\right)$$

por lo tanto, la siguiente desigualdad debe mantenerse para los retornos absolutos para generar mejores resultados que los retornos cuadráticos,

$$\frac{2}{4(\frac{\pi}{2}-1)} \int_0^t \sigma_u^4 du > \int_0^t \sigma_u^2 du \left(\int_0^t \sigma_u du\right)^2$$

No es seguro si la desigualdad se mantendrá o no, por lo tanto se usará una transformación logarítmica donde

$$\log([Y_\delta]_t^{[2]}) - \log\left(\int_0^t \sigma_u^2 du\right) \xrightarrow{L} MN\left(0, \frac{2}{\delta} \frac{\int_0^t \sigma_u^4 du}{\left(\int_0^t \sigma_u^2 du\right)^2}\right)$$

y

$$\log\left(\left([Y_\delta]_t^{[1]}\right)^2\right) - \log\left(\left(\int_0^t \sigma_u du\right)^2\right) \xrightarrow{L} MN\left(0, \frac{4(\frac{\pi}{2}-1)}{\delta} \frac{\int_0^t \sigma_u^2 du}{\left(\int_0^t \sigma_u du\right)^2}\right)$$

para obtener que

$$\frac{2}{4(\frac{\pi}{2}-1)} \int_0^t \sigma_u^4 du > \frac{\left(\int_0^t \sigma_u^2 du\right)^3}{\left(\int_0^t \sigma_u du\right)^2}$$

Nuevamente, no es claro si esta desigualdad es cierta o no. Pero incluso si lo fuera, solamente se podría decir que la variación absoluta realizada cuadrada es un mejor estimador de la volatilidad spot integrada cuadrada que la varianza realizada para la varianza integrada. Sin embargo, este resultado no es exactamente lo buscado, dado que $\left(\int_0^t \sigma_u du\right)^2$ no siempre igualará el objetivo de interés, $\int_0^t \sigma_u^2 du$. Por lo tanto, entonces, se necesita estimar la varianza integrada con $\left(\int_0^t \sigma_u du\right)^2$, agregando un error adicional.

Forsberg y Ghysels (2006) también utilizaron algunos modelos de regresión para predecir ν_{t+1}^2 con ν_t^2 o con $\nu_t^{[1]}$ y obtuvieron un mejor ajuste después de usar $\nu_t^{[1]}$ como regresor. Ghysels, Santa-Clara y Valkanov (2003) obtuvieron resultados empíricos similares.

Aunque no es seguro si la variación absoluta realizada puede generar mejores estimaciones de la varianza integrada que la varianza realizada, los estudios indican el uso de los retornos absolutos en lugar de los retornos cuadráticos. Forsberg y Ghysels (2006) y Andersen, Bollerslev y Diebold (2003) estudiaron la variación bipoder, introducida por Barndorff-Nielsen y Shephard (2004a), como estimador y predictor de la varianza integrada y reportaron que se mejoraría con respecto a la varianza realizada.

4 Conclusiones

Dado un modelo SV para los log-precios y los conceptos de la variación cuadrática y variación elevada a potencia, la disponibilidad de datos en frecuencia alta permitió el uso de una serie de tiempo de variación absoluta realizada para estimar las verdaderas volatilidades. Cuando se usan las variaciones absolutas realizadas crudas como estimador, los errores son altos, especialmente cuando el valor de M fue pequeño. Al utilizar las distribuciones asintóticas de estos errores, las estimaciones se mejoraron mediante un enfoque basado en un modelo y un enfoque sin modelo. Ambos enfoques tendieron a generar resultados similares cuando M tuvo un valor alto y cuando se utilizaron varios componentes rezagados y adelantados en el estimador sin modelo. Aunque, se prefieren los valores absolutos en lugar de valores cuadráticos, la variación absoluta realizada no estima el objeto del interés, es decir, la varianza integrada. Por lo tanto, se tienen que estudiar objetos alternativos como la variación multipoder.

5 Apéndice: Conjunto de datos INTEL

Las acciones INTEL se negocian en la bolsa NASDAQ que se enfoca principalmente en acciones de alta tecnología. El conjunto de datos se construirá con la Base de Datos *Trade and Quotes* (TAQ, por sus siglas en inglés). La base de datos TAQ es la colección de precios de operaciones bursátiles y de cotizaciones intra-día para todos los valores adscritos en todos los mercados de valores principales de los Estados Unidos. Toda la información está disponible en una colección de discos compactos, con 3 a 10 discos separados para cada mes. Más información se encuentra en la página web de la Bolsa de Nueva York (www.nyse.com).

En el presente estudio se trabajará sólo con los precios de transacción, aunque un análisis similar se puede realizar con base en los datos de las cotizaciones. La base de datos TAQ incluye cada precio

de transacción, pero se utilizará el último precio registrado cada cinco minutos para tener datos igualmente espaciados (método de la última cotización instantánea (*tick*), por ejemplo, Wasserfallen y Zimmermann (1985)). Se utilizan los precios de cada cinco minutos desde 9:30 a.m. (cuando abre la bolsa) hasta 4:00 p.m. (cuando cierra la bolsa) para cada día hábil desde el 1 de octubre de 1998 hasta el 29 de septiembre de 2000.

En el conjunto de datos pueden aparecer problemas debido a datos faltantes y divisiones en los títulos accionarios en los mercados (*splits*). El conjunto de datos contiene 39,816 observaciones (79 observaciones cada día para 504 días), y faltan 199 valores. Solamente hay siete días con valores faltantes, y cuatro de estos tienen datos faltantes porque el mercado cerró temprano en estos días. Para obtener una base de datos completa, se puede utilizar una interpolación lineal o puente Browniano. La Gráfica 7a muestra un día con datos faltantes y la Gráfica 7b ilustra ambos métodos (observaciones 56-59 y 61-64 faltan). El problema de interpolación lineal es que la varianza de los retornos será aproximadamente cero para este intervalo, por lo tanto, se utilizará el puente browniano.

Hasta el 24 de junio de 1997, los precios de valores usualmente fueron decretados por la bolsa de Nueva York para que fueran múltiplos enteros de 1/8 de un dólar. Después y hasta el 29 de enero de 2001, fueron múltiplos de 1/16 de un dólar. La comisión cobrada por los agentes de la bolsa fue un número fijo de estas proporciones. Cuando el precio fue demasiado alto, el porcentaje de la comisión fue muy pequeño. La bolsa intentó mantener los precios entre diez y cien, entonces, si el precio fue demasiado alto, la acción se dividió en dos acciones a medio precio. Las divisiones en los títulos accionarios en los mercados (*splits*) no representan un problema para el presente trabajo porque siempre se puede multiplicar la serie por dos a partir del día cuando la acción se dividió. También, se trabajan con precios intra-día y los mercados se dividen al final del día, así no afectan a los retornos.

Gráfica 7. a) Un día con datos faltantes y b) los métodos del Puente Browniano e interpolación lineal para completar las observaciones.

En la Gráfica 8, los precios INTEL se trazan para la muestra completa. Se puede observar cómo los precios se triplicaron a finales de agosto de 2000, pero después los precios disminuyeron abruptamente durante los últimos meses de la muestra. La Gráfica 9a muestra los precios intra-día cada cinco minutos para un día seleccionado aleatoriamente. Con esta estructura de datos,

siempre hay 79 observaciones cada día. En la Gráfica 8b, se pueden observar los retornos cada cinco minutos del conjunto de datos. Los retornos intra-día de cada cinco minutos para el día seleccionado aleatoriamente se muestran en la Gráfica 9b.

Gráfica 8. a) Precios de INTEL sin mercado split, b) retornos a 5 minutos (Intel).

Gráfica 9. a) Precios intra-día cada cinco minutos y b) los correspondientes retornos de la serie de INTEL para un día elegido al azar.

6 Referencias

1. Andersen, T. G. y T. Bollerslev (1997). Intraday periodicity and volatility persistence in Financial markets. *Journal of Empirical Finance* 4, 115-158.
2. Andersen, T. G. y T. Bollerslev (1998a). Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economics Review* 39, 885-905.
3. Andersen, T. G. y T. Bollerslev (1998b). Deutsche mark-dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies. *Journal of Finance* 53, 219-265.
4. Andersen, T. G., T. Bollerslev y F. X. Diebold (2003). Some like it smooth, and some like it rough: untangling continuous and jump components in measuring, modelling and forecasting asset return volatility. Unpublished paper: Economics Department, Duke University.

5. Andersen, T. G., T. Bollerslev y F. X. Diebold (2005). Parametric and nonparametric measurements of volatility. En Y. Aït-Sahalia y L. P. Hansen (Eds.), *Handbook of Financial Econometrics*. Amsterdam: Elsevier Science B.V.
6. Andersen, T. G., T. Bollerslev, F. X. Diebold y H. Ebens (2001). The distribution of realized stock return volatility. *Journal of Financial Economics* 61, 43-76.
7. Andersen, T. G., T. Bollerslev, F. X. Diebold y P. Labys (2001). The distribution of exchange rates volatility. *Journal of American Statistical Association* 96, 42-55.
8. Barndorff-Nielsen, O. E., S. E. Graversen, J. Jacod, M. Podolskij y N. Shephard (2005). A central limit theorem for realised power and bipower variations of continuous semimartingales. En Y. Kabanov y R. Lipster (Eds.), *From Stochastic Analysis to Mathematical Finance*, Festschrift for Albert Shiryaev. New York: Springer Verlag.
9. Barndorff-Nielsen, O. E., S. E. Graversen, J. Jacod y N. Shephard (2006). Limit theorems for bipower variation in financial econometrics. *Econometric Theory* 22, 677-719.
10. Barndorff-Nielsen, O. E., B. Nielsen, N. Shephard y C. Ysusi (2004). Measuring and forecasting Financial variability using realised variance with and without a model. En A. C. Harvey, S. J. Koopman y N. Shephard (Eds.), *State Space and Unobserved Component Models: Theory and Applications*. Proceedings of a Conference in Honour of James Durbin, 205-235. Cambridge: Cambridge University Press.
11. Barndorff-Nielsen, O. E. y N. Shephard (2002). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B* 64, 253-280.
12. Barndorff-Nielsen, O. E. y N. Shephard (2003). Realised power variation and stochastic volatility. *Bernoulli* 9, 243-265.
13. Barndorff-Nielsen O. E. y N. Shephard (2004a). Power and bipower variation with stochastic volatility and jumps (with discussion). *Journal of Financial Econometrics* 2, 1-48.
14. Barndorff-Nielsen, O. E. y N. Shephard (2004b). Econometric analysis of realised covariation: high frequency covariance, regression and correlation in Financial economics. *Econometrica* 72, 885-925.
15. Barndorff-Nielsen O. E. y N. Shephard (2007). Variation, jumps, market frictions and high frequency data in Financial econometrics. En R. Blundell, P. Torsten y W. K. Newey (Eds.), *Advances in Economics and Econometrics. Theory and Applications*, Ninth World Congress. Econometric Society Monographs. Cambridge: Cambridge University Press.
16. Christensen, B. J. y N. R. Prabhala (1998). The relation between implied and realized volatility. *Journal of Financial Economics* 37, 125-150.
17. Comte, F. y E. Renault (1998). Long memory in continuous-time stochastic volatility models. *Mathematical Finance* 8, 291-323.
18. Dacorogna, M. M., R. Gencay, U. A. Müller, R. B. Olsen y O. V. Pictet (2001). *An Introduction to High-Frequency Finance*. San Diego: Academic Press.

19. Ding, Z., C. Granger y R. Engle (1993). A long-memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83-106.
20. Forsberg, L. y E. Ghysels (2006). Why do absolute returns predict volatility so well? Unpublished paper: Department of Economics, University of North Carolina.
21. Ghysels, E., P. Santa-Clara y R. Valkanov (2003). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics*. Próximo a publicarse.
22. Hsieh, D. (1991). Chaos and nonlinear dynamics: Applications to Financial markets. *Journal of Finance* 42, 281-300.
23. Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics* 8, 323-361.
24. Poterba, J. y L. Summers (1986). The persistence of volatility and stock market fluctuations. *American Economic Review* 76, 1124-1141.
25. Protter, P. (1990). *Stochastic Integration and Differential Equations: A New Approach*. New York: Springer.
26. Rosenberg, B. (1972). The behaviour of random variables with nonstationary variance and the distribution of security prices. Working paper 11, Graduate School of Business Administration, University of California, Berkely. Reprinted in Shephard (2005).
27. Schwert, G. W. (1989). Why does stock market volatility change over time? *Journal of Finance* 44, 1115-1153.
28. Shephard, N. (2005). *Stochastic Volatility: Selected Readings*. Oxford: Oxford University Press.
29. Shiryaev, A. N. (1999). *Essentials of Stochastic Finance. Facts, Models, Theory*. Advanced Series on Statistical Science & Applied Probability, Vol. 3. Singapore: World Scientific.
30. Taylor, S. J. (1986). *Modeling Financial Time Series*. New York: Wiley.
31. Taylor, S. J. y X. Xu (1997). The incremental volatility information in one million foreign exchange quotations. *Journal of Empirical Finance* 4, 317-340.
32. Wasserfallen, W. y H. Zimmermann (1985). The behaviour of intraday exchange rates. *Journal of Banking and Finance* 9, 55-72.
33. Zhou, B. (1996). High-frequency data and volatility in foreign exchange rates. *Journal of Business and Economic Statistics* 14, 45-52.