

Por:  
Wilmer O. Martínez R.  
Hector M. Zárate S.

Núm. 665  
2011

# Borradores de **ECONOMÍA**



tá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Col

# Aplicación de las herramientas de control de calidad en la compilación de estadísticas a través del tiempo<sup>1</sup>

Wilmer O. Martínez R.<sup>2</sup>

Hector M. Zárate S.<sup>3</sup>

## Resumen

En este documento se presenta un algoritmo computacional para revisar y controlar eficientemente la calidad de la información de grandes volúmenes de registros administrativos recolectados periódicamente. La metodología utilizada para este propósito es conocida como TOP-DOWN, la cual se basa en un análisis inicial de datos agregados que a su vez determina la verificación de reportes individuales. Esta estrategia se aplica en los reportes de las tasas de interés activas reportadas por el sistema financiero semanalmente a la Superfinanciera y al Banco de la República. El algoritmo es implementado en R.

## Abstract

In this document we implement a methodology known as “TOP-DOWN” to revise and control the quality of historical data with an application to lending interest rates reports. This methodology is based on a systematic review through graphical reports that focus from a macro level (aggregated data for any market segment or account) to micro level. The algorithm is programmed in R.<sup>4</sup>

## Palabras claves:

Control de calidad, TOP-DOWN, series históricas, tasas de interés de colocación.

## Clasificación JEL:

C10, G81, G88

---

<sup>1</sup>“La serie Borradores de Economía es una publicación de la Subgerencia de Estudios Económicos del Banco de la República. Los trabajos son de carácter provisional, las opiniones y posibles errores son responsabilidad exclusiva de los autores y sus contenidos no comprometen al Banco de la República ni a su Junta Directiva”.

<sup>2</sup>*Profesional. Sección Estadística, Banco de la República*

<sup>3</sup>*Econometrista. Unidad de Econometría, Banco de la República*

<sup>4</sup>Agradecemos los comentarios de Eliana González jefe sección Estadística y la colaboración de Deicy Cristiano Profesional sección Estadística.

# 1. Introducción

Una de las funciones de las oficinas de estadística es revisar y asegurar la calidad de los reportes individuales provenientes de diversas fuentes de información del sector financiero y no financiero. Para cumplir con este objetivo, es necesario implementar estrategias eficientes que tengan en cuenta la oportunidad, el gran volumen de datos que periódicamente se recolectan y la creciente demanda de cifras por parte de los usuarios. En este documento se propone una metodología de revisión que tiene en cuenta las series de tiempo de cada uno de los reportes.

Las medidas estadísticas obtenidas periódicamente dependen de la consistencia y oportunidad de los reportes individuales. Por lo anterior, en este documento se propone una metodología basada en la propuesta de Gerrit et al (2008), la cual se fundamenta en la revisión de gráficos de series históricas agregadas en un determinado nivel. Con ésta se detectan patrones atípicos que a su vez son la base para revisar la información desagregada.

Esta metodología se aplica en los reportes semanales de la tasa de colocación que las entidades financieras envían a la Superintendencia Financiera de Colombia (SFC). Dicho enfoque puede ser extendido fácilmente a otros tipos de datos de carácter histórico. Para la revisión sistemática de estos datos en este documento se propone una estrategia basada en gráficos de control de calidad construidos en el programa estadístico R. Para una introducción a este programa se sugiere ver Peter Dalgaard (2002) o Brian S. and Torsten H. (2006).

El resto del documento está organizado de la siguiente manera: en la sección 2 se resumen algunos antecedentes metodológicos utilizados en la revisión de grandes volúmenes de datos y se presenta el nuevo enfoque. En la sección 3 se explica la implementación de la metodología. Una aplicación con los reportes de las tasas de interés activa es implementada en la sección 4 y finalmente en la sección 5 se presentan algunas conclusiones.

## **2. Antecedentes y un cambio de enfoque en la verificación de los datos**

El objetivo principal del control estadístico de calidad es garantizar la confiabilidad de la información agregada, la cual depende principalmente de la consistencia interna de los reportes individuales. Por lo anterior, la identificación de datos atípicos se ha convertido en una labor crucial y las oficinas de estadística han incorporado varias formas de validar los datos. Por ejemplo, con la ayuda de medidas simples, tales como la comparación de tasas de crecimiento con periodos recientes, la revisión del valor más frecuente y su comparación con valores anteriores, el uso de estadísticas descriptivas simples, entre otras, Zárate H. (2011).

Una metodología que utiliza las series a través del tiempo de los reportes individuales e involucra los métodos estadísticos y computacionales apropiados para la detección de datos atípicos, fue implementada por el Banco de España, ver Maravall and Caporello (2002). En ésta, se utilizan de forma automática los modelos ARIMA que detectan y corrigen los datos atípicos y con un pronóstico un periodo adelante se compara con la nueva observación disponible. La regla de decisión se basa en el error de pronóstico; cuando el error de pronóstico es mayor a cierto límite especificado la nueva observación es identificada como un valor atípico, por lo tanto se realiza un proceso de verificación detallado.

### **Propuesta para un cambio de enfoque en la verificación de los datos**

Tradicionalmente, la revisión de los datos se hace con un enfoque del micro-dato, esta manera de proceder, como se caracteriza en Gerrit et al (2008), implica grandes recursos en términos laborales y con un costo de oportunidad de la información considerable. La propuesta de estos autores es considerar un esquema de verificación del nivel macro al micro, minimizando el control de calidad de los individuales y maximizando la calidad y oportunidad de los agregados.

## 2.1. Propuesta de Gerrit et al (2008)

La propuesta de Gerrit et al (2008) permite identificar eficientemente datos atípicos en un nivel macro (para ciertos niveles de agregación). Este método se basa en una inspección visual de los datos agregados e individuales, no solo por medio de tablas sino también por medio de gráficos específicos. Este procedimiento permite identificar posibles problemas a nivel macro y determinar las desagregaciones que deben seguir siendo analizadas.

Dicha propuesta consta de tres etapas: la primera consiste en una inspección visual a un nivel agregado; la segunda utiliza gráficos adecuados para la revisión del micro-dato y la tercera realiza un análisis de consistencia para una validación final.

La primera etapa propuesta por Gerrit et al (2008) considera los siguientes pasos:

- 1) Se calcula la componente de tendencia de la serie de tiempo por medio del filtro de Hodrick Prescott.
- 2) Se calcula la diferencia entre la variable y su componente de tendencia, obteniendo la componente cíclica.
- 3) Se calcula el promedio móvil ( $\bar{X}_t$ ) sobre los periodos  $T - 1$  a  $T - 12$  y la desviación estándar ( $S$ ) histórica hasta la observación  $T - 1$ , para construir el intervalo  $(\bar{X}_t - 2S, \bar{X}_t + 2S)$ .
- 4) Se ajustan los datos atípicos observados en períodos pasados, con el propósito de evitar la sobre estimación del intervalo de control. Así, se revisa que cada una de las diferencias calculadas en el paso 2), se encuentre en el intervalo calculado en 3). Los datos que no se encuentran en el intervalo  $(\bar{X}_t - 2S, \bar{X}_t + 2S)$  se consideran atípicos, para evitar que dichos datos atípicos del pasado enmascaren el análisis del nuevo dato, el ajuste se logra con el promedio de los dos o cuatro vecinos.
- 5) Nuevamente, se calcula el promedio ( $\bar{X}'_t$ ) y la desviación estándar ( $S'$ ), como en 3) con las diferencias corregidas. Posteriormente, el nuevo dato se considera atípico si se encuentra fuera del intervalo  $(\bar{X}'_t - 2S', \bar{X}'_t + 2S')$ .

Es importante mencionar que esta metodología fue desarrollada por Gerit et al (2008) en Excel y automatizada con programación en Visual Basic.

En este documento se diseña el algoritmo y sus respectivos gráficos basado en la primera etapa mencionada anteriormente.

### **3. Implementación de la nueva metodología en la verificación de los datos en tasas de colocación**

En la verificación de datos históricos de tasas de interés, se propone una inspección simultánea únicamente en aquellos casos en que se detecten valores atípicos.

Esta metodología se basa en dos inspecciones visuales para cada nivel de desagregación. La primera considera el uso del *gráfico de control* para el dato más reciente y la segunda evalúa el *comportamiento histórico* de cada serie. A continuación se indica cómo se construye el diagrama del comportamiento histórico y posteriormente el gráfico de control.

#### **3.1. Construcción del diagrama de comportamiento histórico**

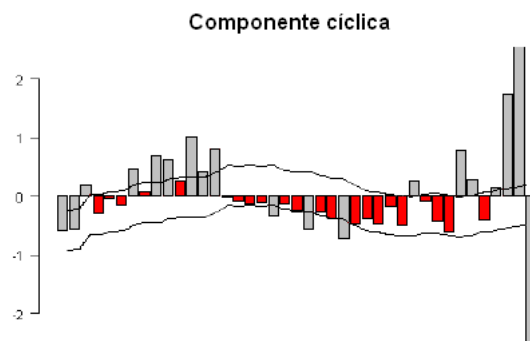
El comportamiento histórico está compuesto por un diagrama en donde el panel superior presenta el gráfico de la serie observada, junto con la componente de tendencia (línea punteada) calculada por medio del filtro de Hodrick Prescott<sup>1</sup> para la variable relevante (montos o tasas de interés). Además, se presenta el valor del reporte anterior (parte inferior) y el valor más reciente (parte superior). En el panel inferior se encuentra un diagrama de barras cuya altura está dada por la componente cíclica (CC) con las correspondientes bandas que permiten determinar si el valor observado se puede clasificar como un valor atípico<sup>2</sup>. Dicho diagrama se compone de barras de distinto color dependiendo si el dato está dentro o fuera de las bandas de control, lo cual es favorable para la inspección (ver figura 1).

---

<sup>1</sup>Con un valor de  $\lambda$  estándar

<sup>2</sup>Las bandas son los límites del intervalo calculado en el numeral 5 de la sección anterior

Figura 1: Diagrama del comportamiento histórico  
**Serie observada y componente de tendencia**



Fuente: Cálculos propios con base en la información del formato 88

La base para la construcción del diagrama de barras sigue la metodología propuesta por Gerrit et al (2008), descrita en la sección 2, con algunas modificaciones en los numerales tres, cuatro y cinco. La primera consiste en calcular el promedio móvil no solamente en el periodo fijo  $T - 1$  a  $T - 12$ , si no en el periodo  $T - 1$  a  $T - \delta$ , donde el usuario puede fijar el periodo más adecuado. Una segunda modificación consiste en calcular el promedio de las diferencias que están bajo control, en el periodo  $T - 1$  a  $T - \delta$ , en lugar del promedio de los dos o cuatro vecinos, como se sugiere en Gerrit et al (2008).

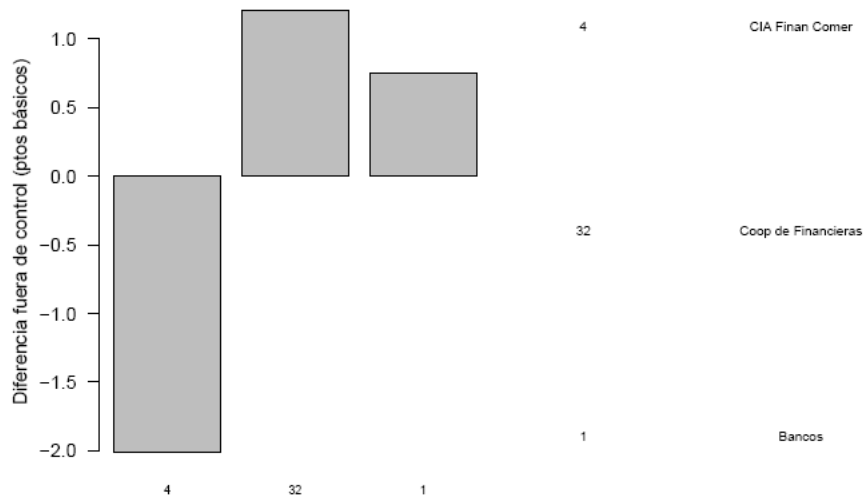
De esta manera, el gráfico permite observar simultáneamente el comportamiento de las series de datos, y así, determinar los valores atípicos. Ahora bien, si se sigue este esquema de chequeo, por ejemplo en el caso de los micro datos, si se cuenta con un conjunto de 53 entidades y si cada una de ellas reporta 10 variables en total se tendrían que revisar 530 gráficos, lo cual sería totalmente ineficiente. Lo ideal es revisar en detalle aquellas variables cuyo último dato se ubique fuera de las bandas control. Para lo anterior se construye un gráfico denominado *gráfico de control*, basado en el último dato del diagrama de barras de la CC.

### 3.2. Gráfico de control

Este gráfico consta de un diagrama de barras del ordenamiento de las diferencias entre el tamaño de la barra del último dato de la Componente Cíclica y la banda de control superior o inferior según sea el caso. Dicha diferencia puede expresarse en puntos básicos o en puntos porcentuales según convenga. En este diagrama se presentan únicamente aquellos casos cuyo último dato exceden éstos límites y por tanto son considerados valores atípicos (ver figura 2). Adicionalmente, cada una de las barras está identificada con el código y la identificación de la cuenta y el sector al cual pertenece. Posteriormente, la revisión se efectúa sobre las series de tiempo cuyo último dato es considerado atípico. La revisión de la información se efectúa del nivel agregado hacia los datos individuales.



Figura 2: Gráfico de control  
**Ordenamiento de agregados atípicos**



Fuente: Cálculos propios con base en la información del formato 88

## 4. Aplicación

Para aplicar la metodología descrita, se consideran los datos de tasas de interés activas (tasas de colocación) que reportan las entidades financieras a la Superintendencia Financiera de Colombia (SFC), y que se reciben en el BR. Estos datos es la base para la generación de reportes estadístico agregado por diferentes clasificaciones. Esta información es reportada semanalmente por las entidades y contiene las distintas cuentas, sectores y tipos de crédito<sup>1</sup>.

### 4.1. Tasas de colocación

Se precisa que las tasas y los montos reportados, por cada entidad o establecimiento de crédito (EC) se refieren al movimiento diario consolidado al último día hábil de cada semana, cabe destacar la responsabilidad sobre la

<sup>1</sup>Estos son: créditos de vivienda, de consumo, microcrédito, comerciales (ordinario, preferencial o corporativo y de tesorería), de tarjetas de crédito, de los descubiertos en cuenta corriente bancaria y de los especiales acordados según las Leyes 550 de 1999 y 617 de 2000; además las entidades pueden ser agrupadas por tipo de entidad (TE), tales como Establecimientos Bancarios, Corporaciones Financieras, Compañías de Financiamiento, Organismos Cooperativos, Entidades especiales y Cooperativas Financieras.

calidad de la información y su oportuna transmisión a la SFC recaerá en el área que determine el representante legal de cada Establecimiento de Crédito.

## **4.2. Recepción de datos**

La información, de las tasas de colocación, es reportada semanalmente. El organismo encargado de recibirla directamente de las entidades financieras es la SFC. Las entidades deben transmitirla a más tardar el segundo día hábil de la semana siguiente a la semana cuya información se está reportando. Luego la SF envía esta información al BR.

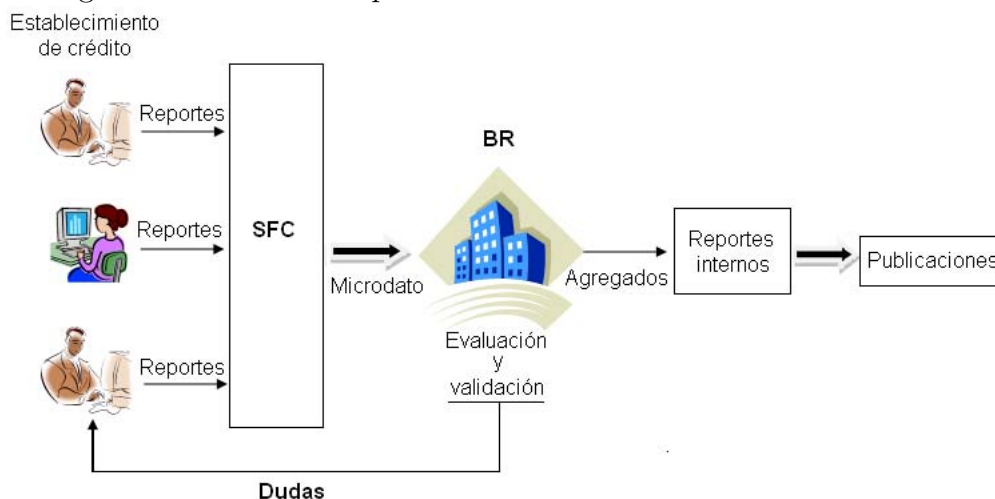
Los reportes individuales deben ser revisados con el fin de verificar la calidad de la información, evitando datos clasificados incorrectamente o que se desvíen notablemente de la situación de mercado. Cuando se detecta algún dato para validar se debe solicitar la confirmación de la entidad.

## **4.3. Proceso de consolidación, verificación de datos y proceso de entrega**

1. Proceso de consolidación. El Administrador de Procesos permite consolidar los datos dentro del sistema, y agiliza la consulta de los datos a través del módulo Consultas en Línea (SISEC). De igual manera, permite hacer consultas por cierto número de semanas, estableciendo previamente las fechas respectivas.
2. Verificación de datos. Para la verificación de los datos es de interés ver el comportamiento de la tasa y el monto agregado por cada TE, así como por cada EC, dentro de cada segmento del mercado, cuenta y plazo. El micro-dato en este caso es la tasa y el monto reportado por cada EC para cada uno de los plazos.
3. Proceso de entrega. El resultado o producto final son los agregados de tasas ponderadas por el monto en determinadas cuentas y segmentos del mercado.

Los anteriores pasos se resumen en el diagrama de la figura 3.

Figura 3: Proceso de captura de información en tasas de colocación



Fuente: Diseño propio

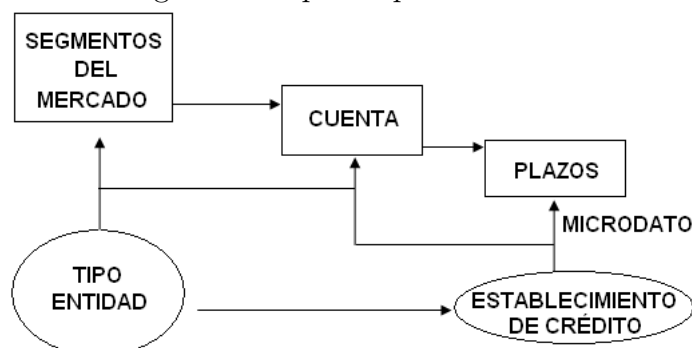
#### 4.4. Aspectos que se deben tener en cuenta en la verificación de los datos

En esta parte del documento se describen los procedimientos actuales para la verificación de las tasas de colocación. De acuerdo con el objetivo de las tasas de interés activas y pasivas la sección de estadística verifica los datos enviados por la SFC, que cada EC reporta a la misma. Cada una de las entidades reporta tanto tasas como montos transados, en cada uno de los plazos estipulados por la SFC. Los aspectos a tener en cuenta en la revisión son los siguientes:

- Si la tasa es mayor a cero el monto también lo sea y en particular que el valor transado no sea exactamente igual al reportado la semana anterior, dado que posiblemente la entidad está repitiendo los mismos datos, esto visto para cada plazo.
- Se verifica que la tasa y el monto no sean muy grandes o muy pequeños a los observados en los últimos períodos.

El sistema de tasas de interés se describe en la figura 4.

Figura 4: Esquema proceso de verificación



Fuente: Diseño propio

#### 4.5. Implementación de la propuesta de revisión

La revisión de los datos depende de la generación de un número considerable de gráficos. Para que ésta sea óptima, se ha diseñado, en el programa R, un menú (ver figura 5), que permite generar reportes gráficos para observar, a diferentes niveles de agregación, el comportamiento de tasas de interés y los montos transados.

Dependiendo de lo observado para cada Tipo de Entidad o para cada Establecimiento de Crédito dentro de algún segmento, cuenta o plazo, el objetivo es establecer si algún EC reportó información errónea o atípica que afecta significativamente a los datos agregados.

El menú está compuesto por opciones que al ser ejecutadas presentan *gráficos de control* para la tasa de interés y el monto transado. En particular, en el diagrama para el monto, en cada barra se señala la participación de cada Tipo de Entidad o Establecimiento de Crédito en el cálculo del agregado de la tasa de interés, para cierta cuenta o plazo (ver figura 6). A partir de lo observado en estos gráficos se decide si es conveniente inspeccionar dicho dato históricamente, en cada caso. En este menú también se tiene la pestaña 'Reporta monto repetido', la cual al ser ejecutada permite detectar las entidades que en los últimos dos reportes los montos son iguales, además se presentan las fechas correspondientes.

La revisión se debe hacer por segmento, cuenta y plazo, del dato agre-

Figura 5: Menú para la consulta de reportes gráficos en tasas de colocación, diseñado en R

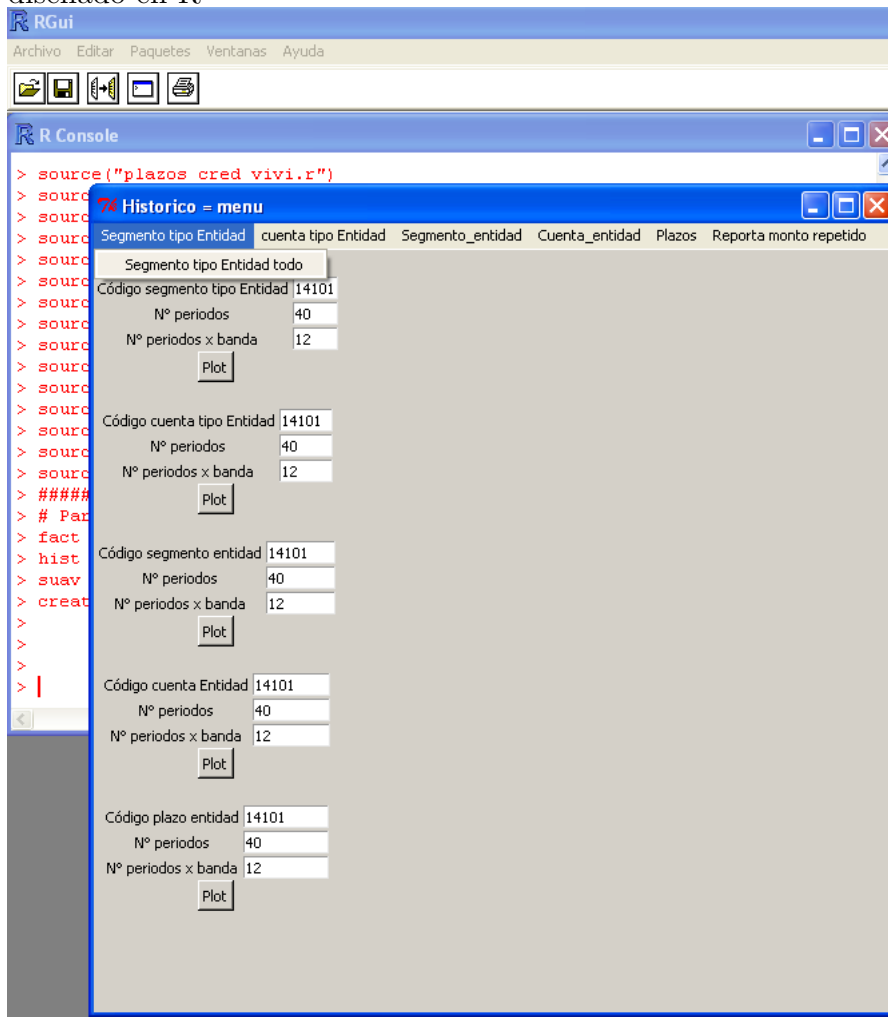
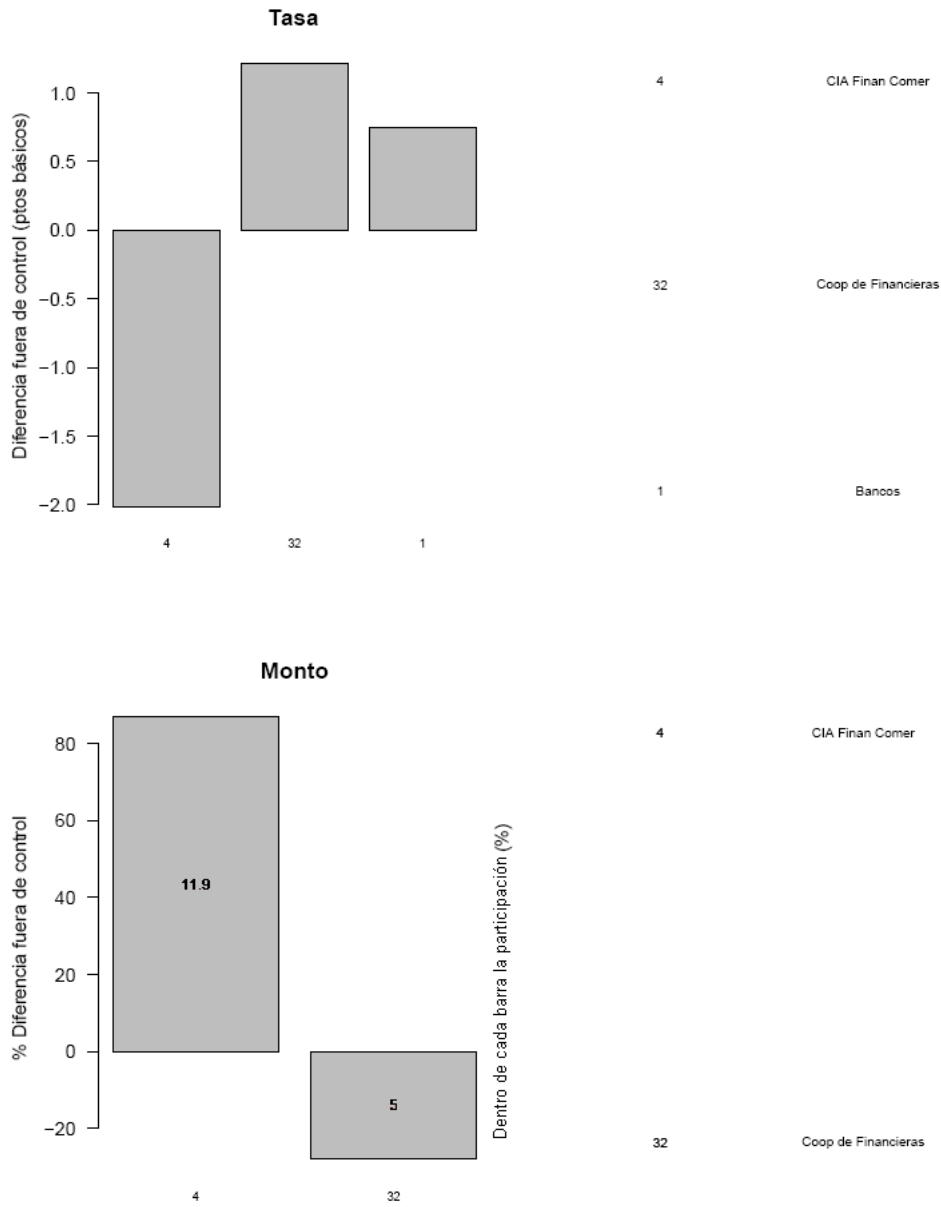


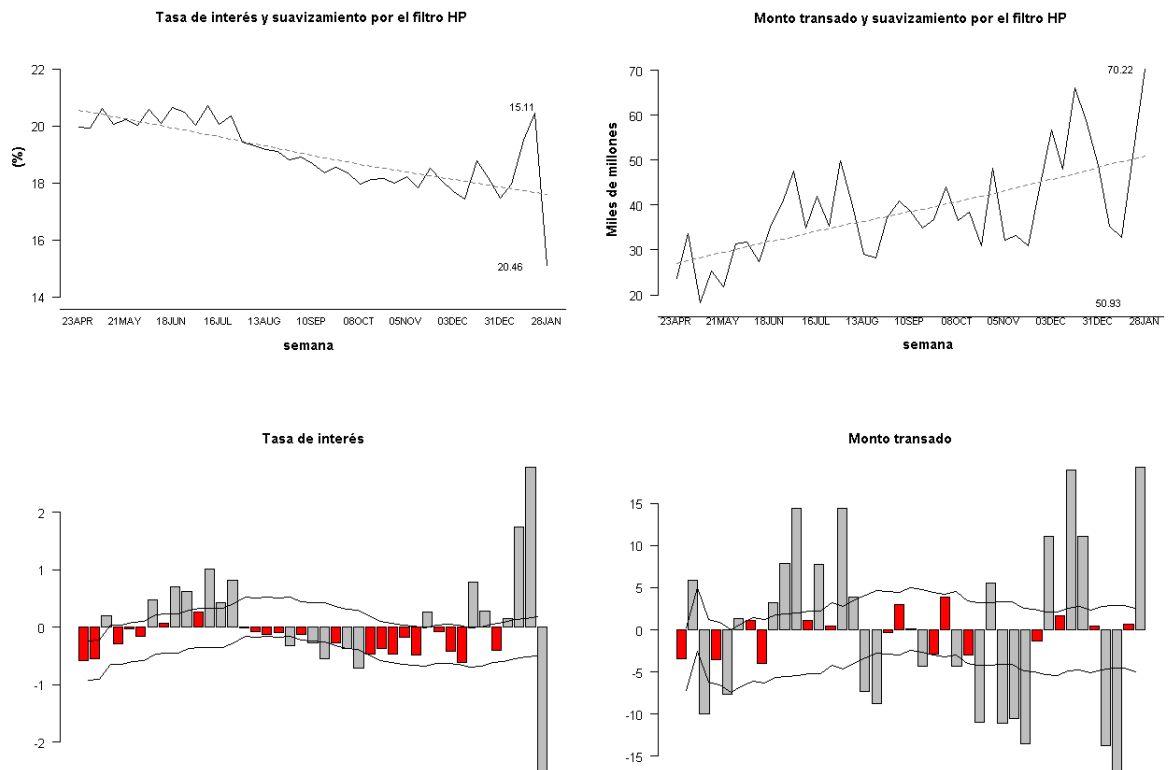
Figura 6: Gráfico de control por tipo de entidad en el segmento, Créditos de Consumo



142 – créditos de consumo

Fuente: Cálculos propios con base en la información del formato 88

Figura 7: Gráfico histórico para el tipo de entidad Compañías de Financiamiento Comercial en Créditos de Consumo



créditos de consumo - CIA Finan Comer

Fuente: Cálculos propios con base en la información del formato 88

gado al dato individual. Inicialmente, se chequean los gráficos de control y en los casos en que sea necesario se revisa el comportamiento histórico de los mismos. A continuación se indica cómo se procede para la inspección de los datos. Para la presente aplicación, el gráfico de control indica si el último dato agregado es un atípico. Esto es, si por ejemplo un tipo de entidad aparece en una de las gráficas se tiene que dicho dato es extremo en tasa o en monto. Luego, la estrategia es inspeccionar estos mismos gráficos por EC y en los casos en que sea conveniente revisar los datos históricos. Los datos históricos pueden ser revisados en cualquier nivel de agregación.

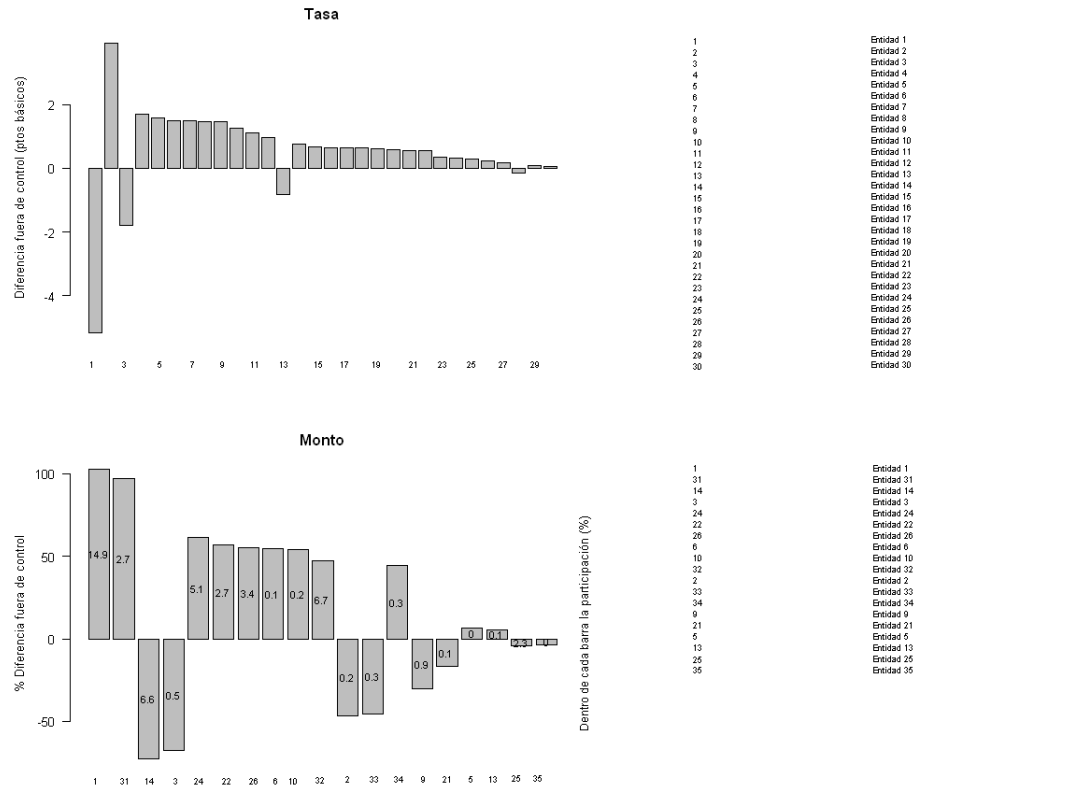
Por ejemplo, en la figura 6, se aprecia que las Compañías de Financiamiento Comercial (CIA Finan Comer) bajaron sus tasas y están fuera de control  $-2$  puntos básicos, aproximadamente. En cuanto a las Cooperativas de Financiamiento (Coop de Financieras) aumentaron sus tasas en más de un punto básico y está por fuera de lo esperado. Por su parte el monto reportado por las Compañías de Financiamiento Comercial, más del 80 % del valor reportado está por fuera de lo esperado, y dicho valor, tiene una participación en el cálculo de la tasa en Créditos de Consumo del 11.9 %. A partir de lo anterior, se sugiere revisar el gráfico histórico para Compañías de Financiamiento Comercial.

Para ello en el menú el campo, “ $N^\circ$  periodos”, hace referencia a la cantidad de semanas que se desean analizar y “ $N^\circ$  períodos  $\times$  banda” es el número de periodos que se usan para calcular las bandas, en otras palabras es el valor  $\delta$  o ventana móvil. Estos campos se sugiere sean los mismos que se utilizaron para la construcción del gráfico de control, luego se obtiene la figura 7. En dicha figura, en tasa se aprecia un cambio de 20.46 puntos básicos a 15.11 y de 50.93 miles de millones de pesos a 70.22 en monto. Posteriormente, se identifican las entidades que están ocasionando el valor extremo.

Para identificar los Establecimientos de Crédito que generan dicha variación, se procede a analizar el segmento específico para cada establecimiento de crédito en cuestión, Figura 8. Adicionalmente se aprecia que la Entidad de Crédito presenta una caída en tasa y un crecimiento en monto y que dicho establecimiento de crédito tiene una participación de 14.9 % en el cálculo de la tasa de este segmento. Por tanto, esta entidad es una de las que más afecta el cambio observado anteriormente. Ahora, si se desea ver el gráfico histórico, se genera la Figura 9 en la cual se aprecia claramente el dato atípico tanto en tasa como en monto. Finalmente, se deben identificar los plazos que



Figura 8: Gráfico de control por Entidad de crédito en el segmento Créditos de Consumo



142 - créditos de consumo

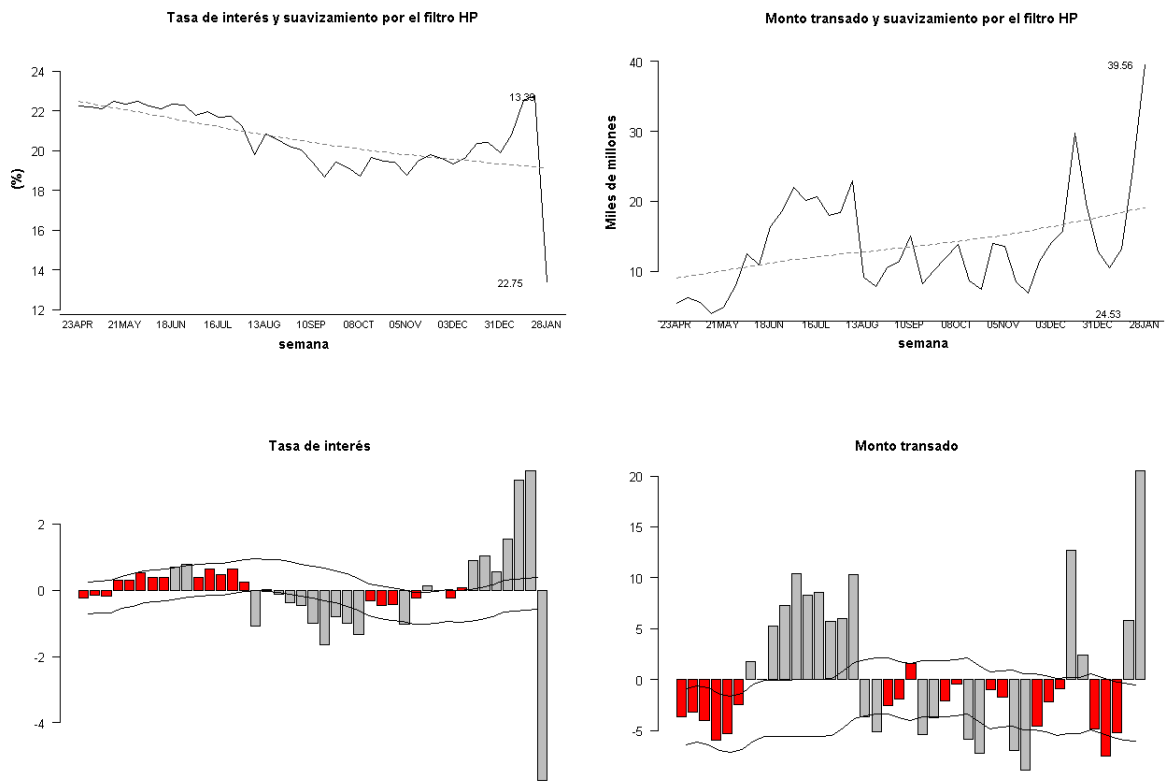
Fuente: Cálculos propios con base en la información del formato 88

ocasionan los valores atípicos. Para ello, el menú ofrece esta opción la cual despliega los gráficos de control para cada uno de los plazos presentes en dicho segmento.

## 5. Conclusiones

La revisión de la calidad de la información es crucial para la generación de reportes con estadísticos agregados. Para tal fin, se propuso una modificación a la metodología de Gerrit et al (2008), la cual permite hacer una

Figura 9: Gráfico histórico para Entidad de crédito en Créditos de Consumo



142xxxxx créditos de consumo - Entidad 1

Fuente: Cálculos propios con base en la información del formato 88

revisión de los datos agregados a los individuales con la metodología (TOP-DOWN). El procedimiento propuesto permite hacer una revisión global de la información en los casos en que sea necesario. En este trabajo se presenta un algoritmo que facilita la revisión y garantiza la confiabilidad de la información, el cual fue implementado en el programa estadístico R. El algoritmo propuesto es flexible y pueda ser adaptable a otros procesos que involucren reportes periódicos de grandes volúmenes de información.

## Referencias bibliográficas

1. Brian S. Everitt and Torsten Hothorn (2006). A Handbook of Statistical Analyses Using R. Chapman & Hall, Boca Raton, FL.
2. Gerrit, van den Dool. and Marc, de Vor. and Dirk van der Wal (2008), ‘A macro-micro approach to compiling statistics’, *Bank for International Settlements* **2**.
3. Maravall, A. and Caporello, G. (2002), ‘A tool for quality control of time series data: programe TERROR’, *Irving Fisher Bulletin*.
4. Peter Dalgaard (2002). Introductory Statistics With R. Springer, New York.
5. Zárate H. (2011). International Encyclopedia of Statistical Science. Banking Statistics, Springer Verlag.