Documentos de Trabajo 2006 | 7

Una metodología bayesiana para promediar predicciones: aplicación al REM del BCRA

Pedro Elosegui / Francisco Lepone / George McCandless
BCRA

Mayo de 2006

ie | BCRA

Investigaciones Económicas
**Banco Central
de la República Argentina**

Banco Central de la República Argentina
ie | Investigaciones Económicas

Mayo, 2006
ISSN 1850-3977
Edición Electrónica

Reconquista 266, C1003ABF
C.A. de Buenos Aires, Argentina
Tel: (5411) 4348-3719/21
Fax: (5411) 4000-1257
Email: investig@bcra.gov.ar
Pag.Web: www.bcra.gov.ar

Una Metodología Bayesiana para Promediar Predicciones: Aplicación al REM del BCRA

Dr. Pedro Elosegui Dr. Francisco Lepone
Dr. George McCandless*
Subgerencia General de Investigaciones Economicas
Banco Central de la República Argentina

May 31, 2006

Abstract

El BCRA publica mensualmente el Relevamiento de Expectativas de Mercado (REM) que resume las proyecciones y predicciones económicas realizadas por un grupo de analistas y consultores económicos. El BCRA da a conocer sólo los principales estadísticos agregados de la muestra, tales como la mediana, el promedio y el desvío estándar. La lógica para usar estos estadísticos es que todos los participantes deben ser ponderados de manera similar.

Si se piensa que algunos consultores poseen mejores modelos subyacentes que otros, la eficacia de los pronosticos agregados puede mejorarse sustancialmente priorizando las predicciones de aquéllos que históricamente han proyectado mejor. Aún desconociendo en detalle los modelos utilizados, se cuenta con la información de las predicciones realizadas por ellos en el pasado. Un método que pondere tal desempeño histórico debería llevar a un mejor promedio agregado. En este trabajo, se desarrolla un método bayesiano que permite calcular tales ponderadores.

El promedio agregado que surge de las ponderaciones bayesianas provee predicciones estadísticamente mejores que la media aritmética, la mediana y otros métodos utilizados usualmente. En particular, el método desarrollado detecta con mayor eficacia cambios de tendencia en las proyecciones.

Las predicciones agregadas publicadas del REM proveen información útil, no sólo para las decisiones de política monetaria y económica, sino también para las decisiones de consumo e inversión, por ende mejorar estas predicciones beneficia a todos los agentes de la economía.

*Agradecemos muy especialmente la colaboración de Francisco Gismondi, Maximiliano Castillo y Diego Ciongo de la Subgerencia General de Economía y Finanzas por la provisión y procesamiento de los datos del REM utilizados en este trabajo. En la elaboración de este trabajo se han respetado todos los términos del convenio de confidencialidad celebrado entre el BCRA y los participantes del REM.

1 Introducción

De manera similar a otros bancos centrales, el BCRA publica mensualmente el Relevamiento de Expectativas de Mercado (REM) que resume las proyecciones y predicciones económicas de corto y mediano plazo realizadas por un grupo de analistas y consultores económicos que participan voluntariamente en el programa. En el marco de la confidencialidad de las proyecciones aportadas por los analistas, el BCRA publica sólo algunos estadísticos agregados de la muestra, tales como la mediana, el promedio y el desvío estándar. Estos estadísticos pueden proveer a las autoridades del banco central de información relevante del consenso profesional sobre la evolución de variables macroeconómicas importantes. Esta información es de utilidad no sólo para la toma de decisiones de política monetaria y económica, sino también para las decisiones de consumo e inversión de los agentes privados de la economía.

Las variables de corto y mediano plazo que se relevan en el REM se agrupan en cinco categorías [BCRA, 2006], índices de precio, variables monetarias y financieras, indicadores de actividad económica, evolución del sector externo y desempeño fiscal. Las proyecciones de corto plazo son encuestadas mensualmente e incluyen proyecciones a uno y dos meses vista. Las variables de mediano plazo son trimestrales y/o anuales, también con proyecciones a dos períodos. Para el caso del PBI nominal y el índice de precios al consumidor se relevan proyecciones a diciembre de cada año. La página de internet del BCRA[1] no sólo publica las principales estadísticas de la muestra, sino también los nombres de las firmas que producen las cinco mejores proyecciones en cada categoría.

La utilidad de la información revelada por el banco central en el marco del convenio de confidencialidad depende críticamente de los estadísticos utilizados para agregar la misma. En este sentido, el objetivo de este trabajo es el de desarrollar una herramienta metodológica que complemente la información estadística publicada actualmente y, a su vez, represente una mejora de pronóstico respecto de otros estadísticos usualmente utilizados (por ej. el promedio de cinco mejores pronósticos publicado por el Banco Central de Brasil). La esencia de la metodología bayesiana radica en la utilización de los errores históricos de cada encuestado para calcular los ponderadores [Canova, 2006; Jacobson, 2004].

Resulta crucial en el potencial éxito del método el supuesto que algunas firmas poseen modelos que son mejores que los disponibles por otras. Bajo esa hipótesis, un método de ponderación que de mayor peso a las predicciones de consultores que han realizado mejores proyecciones en el pasado producirá una mejora en el promedio agregado. Dado que en la práctica no se conoce el modelo individual de cada consultora, debemos aproximar el mismo por la historia de sus predicciones. Las técnicas bayesianas son particularmente apropiadas para explotar esta información y construir tales ponderadores. Sintéticamente, el método consiste en suponer que los errores de predicción poseen una función de verosimilitud con distribución normal de media nula. La varianza de esta distribución constituye un parámetro muy importante que determina cuanto

[1] http://www.bcra.gov.ar

contribuye cada firma al promedio agregado. Si la varianza es pequeña sólo se pondera al mejor pronóstico y si la varianza es muy grande la situación se asemeja a un promedio aritmético simple. El punto es eligir el valor de la varianza de manera de optimizar el uso de la información disponible. Adicionalmente, se discuten otras posibilidades en la elección de varianza así como variantes de los supuestos iniciales (por ejemplo, respecto a la "calidad" inicial asociada a cada consultor).

El trabajo se estructura de la siguiente manera, en la sección 2 se presentan aspectos de la teoría estadística que permiten el cálculo de ponderadores para pronósticos individuales basados en técnicas bayesianas. En la sección 3 se realiza una evaluación del método bajo distintos supuestos iniciales, mostrando algunas de sus propiedades y comparando su desempeño respecto a otras metodologías utilizadas. En la sección 4 se muestra una aplicación de la técnica a datos provenientes del REM. Finalmente, en la sección 5 se reseñan las principales conclusiones y sugieren posibles lineamientos para la generalización del método.

2 El método bayesiano básico

Se asume que cada mes un conjunto A de empresas consultoras envía proyecciones acerca del valor esperado de determinada variable económica al banco central. Llamamos $y_{j,t+1}$ a la proyección realizada a la fecha t por la firma $j \in A$, para tal variable. Se supone que cada firma j posee un modelo (denominado M_j) acerca del cual sólo se conoce la historia de proyecciones $Y_{j,t} = \{y_{j,1}, y_{j,2}, ...y_{j,t}\}$ Estamos interesados en obtener el mejor promedio ponderado de las proyecciones correspondientes a un período. La información disponible es la historia de las realizaciones de la variable en cuestión $y^t = \{y_1, y_2, ...y_t\}$ donde y_t son las realizaciones de la variable a la fecha t. Dado que se cuenta con la historia de predicciones de cada firma y la realización de la variable y^t, es posible construir una serie de errores de predicción históricos para cada firma participante: $\varepsilon_{j,t} = y_t - y_{j,t}$ y $\Gamma_{j,t} = \{\varepsilon_{j,1}, \varepsilon_{j,1}, ...\varepsilon_{j,t}\}$.

Queremos calcular un promedio agregado de la forma:

$$E_t\left(y_{t+1}\right) = \sum_{j\in A} E_t\left(y_{j,t+1}|M_j, y^t\right) g\left(M_j|y^t\right),$$

donde $E_t\left(y_{j,t+1}|M_j, y^t\right)$ es la predicción que realiza la firma j en el período t sobre el valor de la variable en $t+1$ (que se supone coincide con el valor que envían al Banco Central). El valor $g\left(M_j|y^t\right)$, es la ponderación o *posterior* que se asigna a la predicción de la firma j. El *posterior* utilizado para la predicción del modelo de la firma j estará dado por:

$$g\left(M_j|y^t\right) = \frac{f\left(y^t|M_j\right) g\left(M_j\right)}{f\left(y^t\right)}$$

Donde $g\left(M_j\right)$ constituye el supuesto a priori *(prior)* acerca de la calidad relativa de cada modelo y $f\left(y^t|M_j\right)$ es la función de verosimilitud del modelo j

dados los datos observados. Esta función es también denominada probabilidad condicional conjunta de los datos dado el modelo. En tanto, $f(y^t)$ es la probabilidad no condicional de los datos. Por último, la ponderación bayesiana o *posterior* $g(M_j|y^t)$ es también denominada probabilidad del modelo condicional a los datos observados.[2] Con un conjunto finito de modelos, A, la probabilidad de los datos es simplemente,

$$f\left(y^t\right) = \sum_{j \in A} f\left(y^t|M_j\right) g\left(M_j\right).$$

2.1 Supuesto de *prior* uniforme

Empezamos suponiendo un *prior* uniforme, es decir no informativo acerca de la calidad relativa de los modelos subyacentes en las proyecciones de las firmas consultoras. De manera que si hay n firmas en A, entonces $g(M_j) = 1/n$ (promedio simple). Para la función de verosimilitud, $f(y^t|M_j)$ se asume una distribución normal con media cero, que esta dada por los errores de predicción de los modelos y la diferencia entre tales proyecciones y las realizaciones observadas de las variables. Inicialmente, se supone una varianza σ^2 constante, cuya magnitud es sumamente relevante ya que, como se muestra con mayor detalle en secciones subsiguientes, modificaciones en su valor cambian las ponderaciones relativas de los diferentes modelos. Dados estos supuestos, si se cuenta con T observaciones de los errores de predicción de las firmas j, la función de verosimilitud $f(y^t|M_j)$ esta representada por

$$f\left(y^T|M_j\right) = [2\pi\sigma]^{-.5T} \exp\left[-.5\frac{\sum_{t=1}^{T}\left(\varepsilon_{j,t}-0\right)^2}{\sigma^2}\right].$$

Bajo la hipótesis de *prior* no informativa, la ponderación asociada a la predicción de la firma j será

$$\frac{f\left(y^T|M_j\right)}{\sum_{k\in A} f\left(y^T|M_k\right)}.$$

donde, como corresponde de la definición, las ponderaciones suman la unidad.

2.2 Un *prior* no uniforme

El desarrollo anterior se realizó suponiendo un *prior* uniforme para todos los modelos, a partir de tal supuesto inicial las ponderaciones posteriores bayesianas (*posteriors*), como se muestra más adelante, permiten mejorar el promedio agregado a partir de la información aportada por los datos históricos. Si la precisión de las predicciones de las firmas tiene algún nivel de correlación entre variables, la información de los *posteriors* correspondiente a una variable i^* secundaria

[2] La distribución de probabilidad conjunta $f(x,y)$ es igual a la probabilidad condicional, $f(x|y)$ o $f(y|x)$, multiplicada por la respetiva probabilidad marginal, $g(y)$ o $g(x)$. Esto da, $f(x,y) = f(x|y)g(y) = f(y|x)g(x)$ que puede ser reescrito como la denominada *Regla de Bayes*, $f(y|x) = \frac{f(x|y)g(y)}{g(x)}$.

puede ser utilizada como *prior* para mejorar las ponderaciones del promedio bayesiano de otra variable i de mayor interés. Si tales *posteriors* contienen información relevante al ser utilizadas como *prior* podrían mejorar las ponderaciones posteriores de la segunda variable en relación a un supuesto de *prior* uniforme de la misma.

El supuesto básico en este caso, es que algunas firmas poseen modelos de predicción que son en general mejores que los correspondientes a otras firmas. Este supuesto implica que debe haber alguna correlación entre los errores de predicción: las firmas que tienen menores errores de predicción en una variable tenderán a presentar menores errores en otras variables. De ser cierta, esta característica, podría ser explotada y utilizada como *prior* informada o no uniforme.

Suponga que el Banco Central esta interesado en predecir la inflación y la tasa de interés, normalmente las variables de principal interés, particularmente en el caso de bancos centrales que utilizan metas de inflación como su objetivo de política. En tal caso, como alternativa a utilizar un *prior* uniforme para estas variables, podrían utilizarse datos correspondientes a variables reales, tales como el producto, para calcular *posteriors* correspondientes a cada firma y utilizar los mismos como *prior* para la estimación de las ponderaciones bayesianas correspondientes a la inflación y la tasa de interés. Los promedios bayesianos de estas variables mejorarían, en relación a los realizados con un *prior* uniforme, en la medida que la información contenida en los ponderadores correspondientes a la variable real (producto) sea un buen predictor de la habilidad de esas firmas para predecir la inflación o la tasa de interés. Debe tenerse en cuenta que la predicción de algunas variables usualmente (formal o informalmente) involucran el uso de la predicción de otras variables (por ejemplo, la predicción de la tasa de interés nominal requiere el uso de la tasa de inflación esperada), por lo tanto estas variables no son apropiadas para utilizarse como *prior* una de la otra, ya que se estaría utilizando practicamente la misma información.

Entonces, para encontrar un mejor esquema de *posteriors*, utilizamos el supuesto de que existe correlación entre la precisión de las predicciones de las diferentes variables para las firmas. Esto es diferente de la correlación entre las predicciones que una firma realiza de las variables (tal como la que pueda surgir entre la predicción de la inflación y la tasa de interés nominal) o la correlación entre las variables mismas. Especificamente, sean $y^{t,i} = \left\{y_1^i, y_2^i, ...y_t^i\right\}$ las realizaciones de la variable i y sea $Y_{j,t}^i = \left\{y_{j,0}^i, y_{j,1}^i, ...y_{j,t}^i\right\}$ el conjunto de predicciones de la variable i realizado por la firma $j \in A$. Se elige una variable i^*, y se calculan los pesos posteriores para cada firma j utilizando

$$g\left(M_j|y^{t,i^*}\right) = \frac{f\left(y^{t,i^*}|M_j\right) g\left(M_j\right)}{f\left(y^{t,i^*}\right)},$$

donde

$$f\left(y^{t,i^*}\right) = \sum_{j \in A} f\left(y^{t,i^*}|M_j\right) g\left(M_j\right).$$

Ahora utilizando los *posteriors* $g\left(M_j|y^{t,i^*}\right)$ como *prior* para otra variable $i \neq i^*$,

	M_1	M_2	M_3	M_4	M_5
$\boldsymbol{\sigma}_i$	0.284	0.197	0.292	0.184	0.287
	M_6	M_7	M_8	M_9	M_{10}
$\boldsymbol{\sigma}_i$	0.269	0.218	0.221	0.334	0.147

Table 1: Desvío estándar correspondiente a cada modelo.

se calculan *posteriors* "mejorados" como

$$g\left(M_j|y^{t,i}\right) = \frac{f\left(y^{t,i}|M_j\right) g\left(M_j|y^{t,i^*}\right)}{f\left(y^{t,i}\right)},$$

donde

$$f\left(y^{t,i}\right) = \sum_{j \in A} f\left(y^{t,i}|M_j\right) g\left(M_j|y^{t,i^*}\right).$$

Este nuevo *prior* es informativo, con información proveniente de la habilidad de las firmas de predecir alguna otra variable.

3 Evaluación del método propuesto

A fin de evaluar los potenciales beneficios del método de promedios bayesianos se utilizan datos de predicciones generadas artificialmente, para comparar el desempeño del mismo respecto a otros cinco métodos utilizados usualmente para computar proyecciones agregadas. Los métodos alternativos utilizados son el promedio aritmético simple de todos los modelos, el promedio aritmético simple de los cinco modelos de mejor predicción, la mediana de la distribución de predicciones y el método de combinación de pronósticos propuesto por Granger y Bates (1969). El conjunto de información generado artificialmente esta compuesto por 10 firmas consultoras, que se supone utilizan modelos para realizar predicciones acerca de una variable cuya realización es constante e igual a uno. Los pronósticos de la firma i tienen una distribución $N(1, \sigma_i^2)$, donde los σ_i^2 se distribuyen uniformemente entre $[0, 0.25]$. Estos modelos subyacentes son muy simples, cada predicción de una firma se genera a partir de una distribución normal con varianzas diferentes entre firmas. En cada simulación se utilizan 101 observaciones para cada una de las 10 firmas. Las primeras 100 observaciones se emplean para calcular las ponderaciones (en el método bayesiano y en el método de promedio agregado) y la observación 101 se utiliza como proyección. Para el ejemplo, en la Tabla 1 se muestran los errores estándar σ_i.

La Figura 1 muestra las proyecciones generadas a partir de las muestras distribuidas normalmente (el modelo) de cada firma. Dado que el dato predicho por cada modelo es una constante igual a la media de la distribución de predicciones de cada firma, los modelos M_2, M_4 y M_{10} son mejores que el resto ya que tienen un menor σ_i asociado.

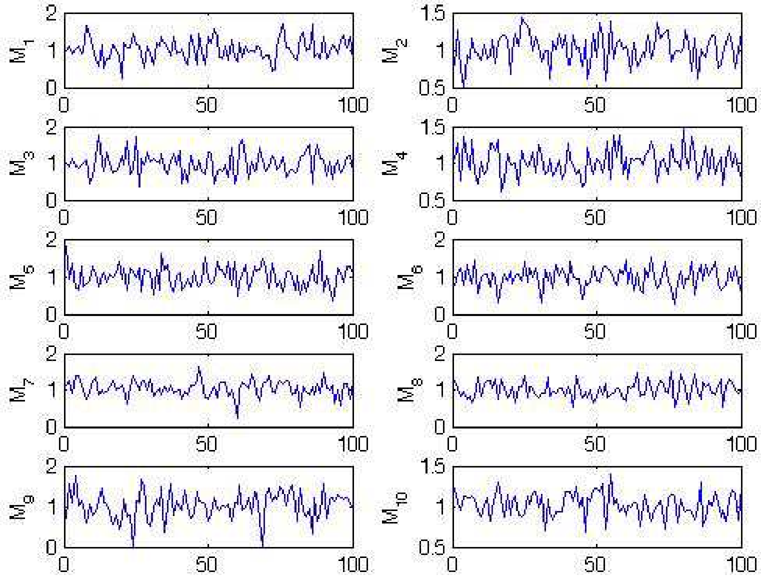

Figure 1: Pronósticos asociados a cada modelo.

Para el promedio bayesiano, elegimos un *prior* sobre los diferentes modelos: uno que da la misma probabilidad ($1/k$ donde $k = 10$ es el número de firmas) a la predicción de cada firma. Este prior asume que los pronósticos de cada firma tiene la misma calidad, más adelante se analizará este supuesto con más detalle.

Como fue mencionado en las secciones previas, es necesario asignar un valor al desvío estándar σ_0 de la función de verosimilitud. Del análisis surge que σ_0 es un parámetro crucial en la determinación de las ponderaciones bayesianas (y, consecuentemente, es importante en la determinación de la eficiencia del pronóstico agregado). La razón de esto es que σ_0 controla tanto la altura como la amplitud de la distribución. Un valor de σ_0 que es demasiado grande da como resultado errores de proyección con probabilidades uniformes y la función de verosimilitud resultante de los diferentes modelos se vuelven practicamente indistinguibles. Un valor de σ_0 demasiado pequeño será restrictivo y ponderará unicamente el modelo con menores errores, ignorando los restantes. Considerando esto, se propone emplear el valor de σ_{opt} para el cual los errores cuadráticos históricos (los 100 primeros datos) de las proyecciones bayesianas son mínimos. La Figura 2 muestra los errores cuadráticos de predicciones bayesianas como función de σ_0 generados a partir de la muestra simulada.

La minimización de esta función arroja $\sigma_{opt} = 0.5933$ para el presente ejemplo. Con este valor de σ_{opt} es posible calcular las ponderaciones bayesianas de

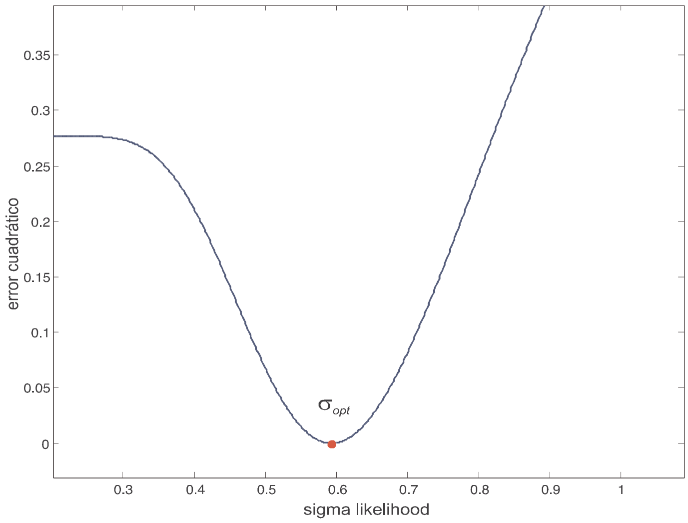

Figure 2: Error cuadrático acumulado como función de σ_0.

wb_1	wb_2	wb_3	wb_4	wb_5
2.0×10^{-4}	0.0685	1.0×10^{-4}	0.1373	1.61×10^{-4}
wb_6	wb_7	wb_8	wb_9	wb_{10}
5.91×10^{-4}	0.016	0.0174	2.40×10^{-6}	0.76

Table 2: Ponderaciones bayesianas para sigma óptimo.

los modelos. Estas se muestran en la Tabla 2.

Hacemos notar aquí que el hecho que $wb_7 < wb_8$ cuando $\boldsymbol{\sigma}_{M_7} < \boldsymbol{\sigma}_{M_8}$ se debe a que en el calculo de ponderadores se asume que los errores poseen media cero mientras que las varianzas muestrales se calculan respecto a la media de cada conjunto de datos.

Una vez obtenidos los ponderadores, es directo calcular el promedio bayesiano de las predicciones, $(\overline{y}_b)$. Como se mencionara previamente, también se incluyen otras metodologías a fines comparativos: 1) un promedio aritmético simple de todas las empresas, 2) el método de combinación de pronósticos[3], 3) la predicción dada por el mejor modelo [4], 4) un promedio aritmético de las firmas "top 5", y 5) la mediana de la distribución de todas las predicciones. En la Tabla 3 se muestran los resultados de cada método.

Tal como puede observarse en la Tabla 3 los pronósticos de la media aritmética y de varianza mínima del ejemplo tienen errores de $\sim 8.8\%$ y $\sim 10.8\%$ respectivamente, en tanto que la proyección bayesiana resulta exacta. La Figura

[3] Sintéticamente, este método elige ponderadores que minimizan la varianza de los errores de predicciones pasadas.

[4] Se entiende como el mejor modelo aquel de mínima varianza.

Método	Predicción
Prom. Bayesiano	1.00
Prom. Aritmético	0.892
Prom. Min. Varianza	0.912
Mejor modelo	1.053
"Top 5"	0.883
Mediana	0.924
Valor real	1.00

Table 3: Pronósticos obtenidos con distintos métodos.

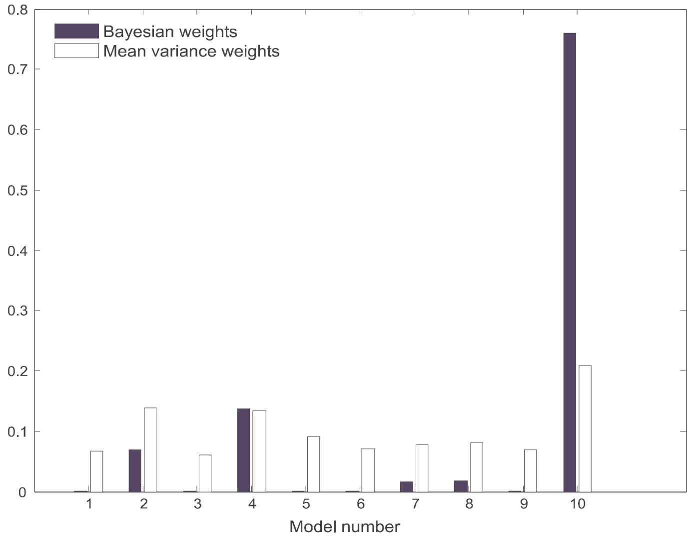

Figure 3: Ponderaciones correspondientes a los métodos bayesiano y de varianza mínima.

3, muestra las ponderaciones bayesianas y de varianza mínima, correspondientes al ejemplo particular bajo análisis.

A fin de comparar la eficiencia relativa de los seis métodos, se desarrolló un conjunto de *tests*. Cada *test* consistió en diez mil ejercicios independientes de proyección con las características básicas del ejemplo descripto (la varianza de las 10 en cada test se eligió de una distribución uniforme con parámetros $[\sigma_a^2, \sigma_b^2]$, donde los límites σ_a^2 y σ_b^2 se tomaron de manera distinta en cada test). Los resultados que se muestran a en la Tabla 4 corresponden al intervalo $\sigma_a^2 = 0.01, \sigma_b^2 = 0.25$.[5] Allí, en el 19.0% de los casos el promedio ponderado bayesiano dio la mejor predicción. En tanto, el mejor modelo fue primero en el 16.35% de los casos, el promedio de varianza mínima en el 18.1%, el promedio aritmético de

[5] En el Apéndice 7 se detallan los resultados para otros valores de σ_a^2 y σ_b^2.

Método	% error de proyección	Std.	Indice	Ranking
Prom. Bayesiano	4.73%	0.0620	2.946	1
Prom. Aritmético	7.14%	0.0903	3.958	6
Prom. Min. Varianza	4.91%	0.0645	3.035	2
Mejor modelo	7.02%	0.0964	3.679	4
"Top 5"	6.41%	0.0830	3.655	3
Mediana	6.39%	0.0827	3.720	5

Table 4: Ranking de eficiencia.

los "top 5" el 15.9% de las veces, la mediana el 16.5%, y el promedio aritmético simple el restante 14.15%.

Es posible evaluar el desempeño global de los diferentes métodos considerados, a travé de distintos criterios entre los que consideramos tres 1) el promedio del error porcentual absoluto de las predicciones, 2) el desvío estándar y 3) un índice de desempeño[6] que posee un rango que va de 1 (mejor) a 6 (peor). La Tabla 4 muestra los resultados junto con las posiciones relativas de los diferentes estadísticos.

De la Tabla 4 surge que el promedio bayesiano es el método de mejor ajuste, seguido por el método de mínima varianza y por el mejor modelo. Analizando los resultados obtenidos para otros intervalos de varianzas (Apéndice 7) surge que cuando el límite inferior $\sigma_a^2 = 0$ el mejor método es el de mínima varianza, en el límite opuesto $\sigma_a^2 \rightarrow \sigma_b^2$ el método de mejor desempeño es el promedio aritmético simple y en todos los casos intermedios el método bayesiano es el que se impone. Adicionalmente, es el de mejor comportamiento global para para el rango $[\sigma_a^2, \sigma_b^2]$ estudiado.

Es interesante considerar la sensibilidad de los diferentes métodos a la longitud del conjunto de información. A tal fin se analiza el desempeño relativo de los modelos como función del número de predicciones pasadas ($n_t = 5, 15, 20, 100$). Los resultados muestran que cuando la muestra es pequena el método bayesiano es más eficiente que las restantes alternativas (Tabla 5). Como podría esperarse, cuando la muestra posee una historia suficientemente grande el método de mínima varianza es el que se impone.

Otra cuestión interesante a considerar es la relación entre el número de datos históricos y la estabilidad de los ponderadores. Para analizar esta relación se parte de una corrida aleatoria común, se genera un conjunto de proyecciones de diferente longitud y se computan las respectivas ponderaciones bayesianas. Los resultados se muestran en la Figura 4. Puede verse que, si existe poca informa-

[6] El índice se computa de la siguiente manera. La posición relativa en el ranking de cada ejercicio de predicción es puntuada desde 1 a 6. Entonces la posición en el ranking de cada modelo M_i se calcula así:

$$\sum_k p_{ik} \times k$$

donde p_{ik} es la probabilidad del modelo i de tener un ranking k. Obviamente, menor el valor del índice mejor el desempeño del modelo.

Método	$n_t = 5$	$n_t = 15$	$n_t = 20$	$n_t = 100$
Prom. Bayesiano	2.86	2.78	2.89	2.97
Prom. Aritmético	3.47	3.74	3.78	3.93
Prom. Min. Varianza	4.73	4.01	3.44	2.95
Mejor modelo	3.50	3.44	3.66	3.87
"Top 5"	3.38	3.55	4.64	3.59
Mediana	3.05	3.48	4.60	3.70

Table 5: Indice de eficiencia en función de la cantidad de datos disponibles.

ción previa, ésta no es demasiado informativa y las ponderaciones bayesianas cambian a medida que se dispone de mayor cantidad de información. Cuando la cantidad de información histórica crece las ponderaciones se estabilizan.

De la Figura 4 puede concluirse que se necesitan entre 20 y 30 observaciones históricas para que las ponderaciones alcancen valores similares a sus valores asintóticos.

A fin de evaluar la importancia de utilizar o no *priors* informativos, se diseño el siguiente ejercicio: 1) a partir de una muestra inicial de 40 datos y utilizando *priors* uniformes, se hallaron *posteriors* para ser utilizados como *priors* en los pasos siguientes, 2) dados distintos valores de n, se generaron diez mil conjuntos de prediccciones independientes, 3) para cada conjunto se obtuvieron *posteriors* utilizando los *priors* calculados en el paso 1, 4) se calcularon las ponderaciones bayesianas utilizando los dos conjuntos de *posteriors* obtenidos en 2 y en 4 y se evaluó el desempeño relativo en predecir los datos. La Tabla 6 muestra los resultados. Como es esperable el tener un *prior* razonable mejora los resultados, especialmente en el caso de muestras pequeñas. Por ejemplo, para una muestra de sólo 20 datos, la eficiencia sobre un supuesto de *prior* uniforme aumenta en un factor de 2.65. A medida que el número de observaciones históricas aumenta, el *prior* se vuelve menos importante. Es interesante notar que el uso de un *prior* sesgado y poco informativo puede ser perjudicial, especialmente si se cuenta con suficiente información histórica. Por caso, si se cuentan con 100 o más datos históricos, el uso de un *prior* no uniforme puede ser perjudicial, ya que podría agregar un sesgo, cuando ya se cuenta con información suficiente para generar *posteriors* estables.

3.1 Elección de la varianza en la función de verosimilitud

Es necesario realizar algunos supuestos acerca de las mega propiedades del modelo. Entre esas propiedades se encuentra el supuesto acerca de la forma de la función de verosimilitud común, $f\left(y^t|M_j\right)$, que se utiliza en la ecuación que da como resultado las ponderaciones a aplicar en cada modelo.[7]

$$g\left(M_j|y^t\right) = \frac{f\left(y^t|M_j\right) g\left(M_j\right)}{f\left(y^t\right)}$$

[7]Recuerde que lo único que se conoce de cada modelo es el conjunto de pronósticos que han enviado al banco central. Los pronósticos son los modelos.

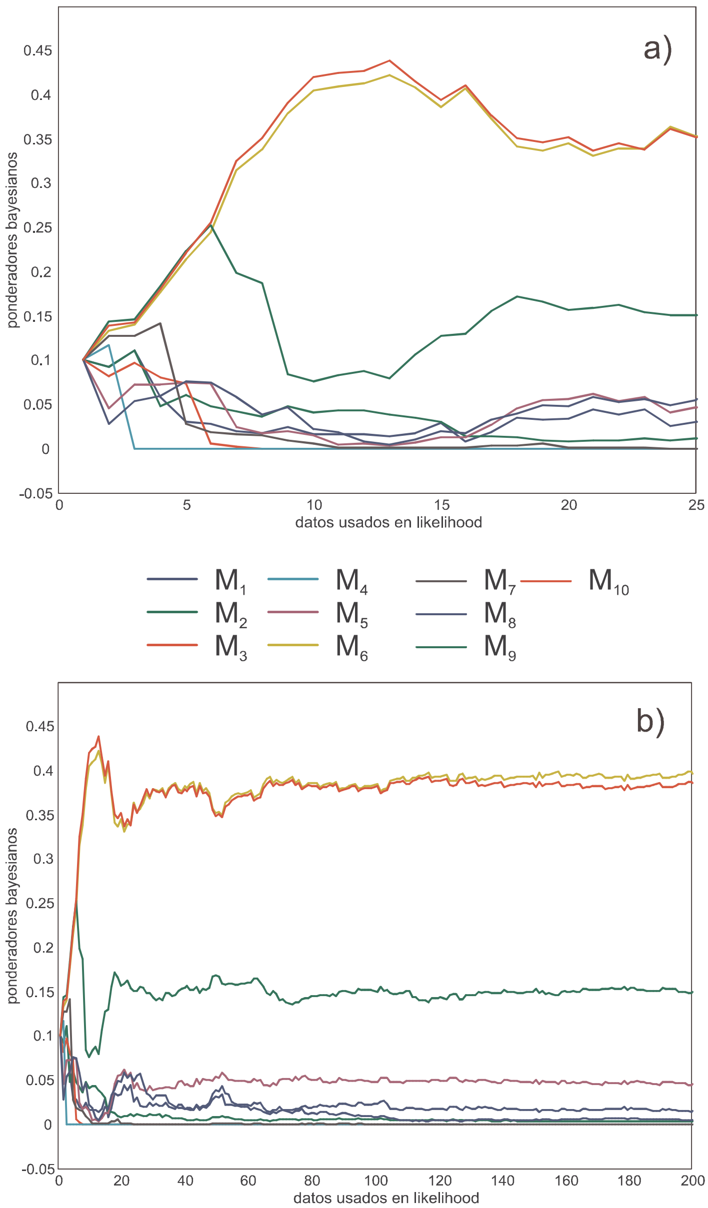

Figure 4: Estabilidad de los ponderadores bayesianos como función de la cantidad de datos históricos disponibles.

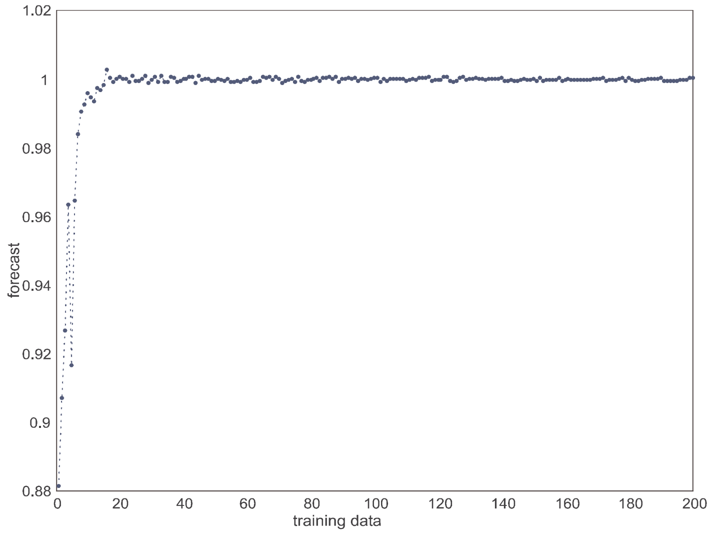

Figure 5: Pronóstico bayesiano como función del número de datos históricos.

n	Prior informativo	Prior constante
3	85.7%	14.3%
5	80.9%	19.1%
10	76.5%	23.5%
20	72.6%	27.4%
30	64.8%	35.2%
40	63.4%	36.6%
50	60.6%	39.4%
60	61.7%	38.3%
100	53.1%	46.9%
200	49.85%	50.15%
500	48.0%	52.0%

Table 6: Eficiencia relativa entre el uso de prior uniforme e informativa como función de la cantidad de información disponible.

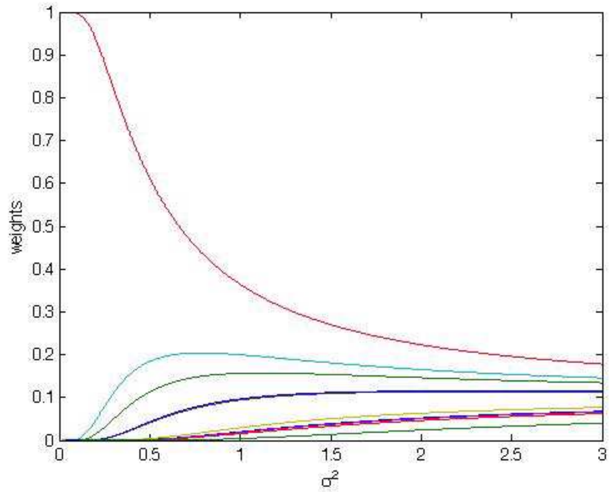

Figure 6: Ponderadores bayesianos en función de σ^2.

Se supone que los errores están normalmente distribuidos, con media cero cuando se mide en término de los errores entre las predicciones de los modelos y los datos observados, donde la varianza es la misma entre los modelos. Dado estos supuestos, la función de verosimilitud debería escribirse en realidad como,

$$f\left(y^t | M_j, \sigma^2, \mu = 0\right).$$

La distribución posterior y las ponderaciones resultantes dependen de la elección de σ^2. Por esta razón, sería mejor expresar las ponderaciones condicionales en los mega-parámetros como

$$g\left(M_j | y^t, \sigma^2, \mu = 0\right)$$

Para los datos artificiales utilizados en el trabajo, la Figura 6 muestra las ponderaciones calculadas para un rango entre 0.01 y 3 de la varianza de la función de verosimilitud. Cada línea en la figura muestra las ponderaciones asignadas a un modelo determinado a medida que cambia el valor de σ^2. Como fuera mencionado, cuando σ^2 es muy pequeño, solo el mejor modelo recibe ponderación positiva. A medida que σ^2 aumenta, más modelos reciben ponderación positiva, hasta que en el límite cuando σ^2 se acerca a infinito todos los modelos reciben igual ponderación. La elección del σ^2 óptimo es un ejercicio que no tiene una respuesta obvia.

El método propuesto resuelve un problema de optimización bien definido. Se utilizan las predicciones pasadas y la información disponible para encontrar el $\widehat{\sigma}^2$ que haga que las ponderaciones minimizen el cuadrado de los errores de las predicciones agregadas realizadas utilizando las ponderaciones y las predicciones de cada modelo por separado. Esta metodología se condice con los objetivos del ejercicio, que precisamente consiste en encontrar las ponderaciones óptimas para tales predicciones.

Para el conjunto de observaciones hasta el período t, y^t, las predicciones pasadas de esta variable de parte de las Q firmas participantes, $\{y_{j,s-1}\}_{s=0}^{t-1}$, y un valor de σ^2 perteneciente al dominio permitido, se obtienen las ponderaciones $w_{j,t} = g\left(M_j|y^t, \sigma^2, \mu = 0\right)$, para $j = 1, ..., Q$. Para este valor de σ^2, puede encontrarse los ponderadores ponderados en la muestra,

$$E_t\left(y_{s+1}\right) = \sum_{j=1}^{Q} w_{j,t} y_{j,s},$$

para $s = 0, ..., t-1$, donde $E_t\left(y_{s+1}\right)$ se utiliza para indicar las predicciones ponderadas en la muestra utilizando predicciones históricas de las firmas pero promediadas utilizando pesos que fueron encontrados utilizando todos los datos disponibles hasta el período t. Buscamos los valores de σ^2 que minimizan los errores cuadráticos,

$$\sum_{s=1}^{T}\left[y_{s+1} - \sum_{j=1}^{Q} w_{j,T} y_{j,s}\right]^2.$$

Para los datos simulados, los valores de esta función para un rango de σ^2 entre .04 y.81 (con σ de.2 a .9) se muestran en la Figure 2. El error cuadrático mínimo ocurre en $\widehat{\sigma}^2 = .352$ (o $\widehat{\sigma} = .5933$).

Técnicamente este problema podría resolverse de manera diferente. en particular, un enfoque alternativo puede surgir a partir de utilizar la propia distribución de la varianza de los datos como un método potencial de ponderación. Esto es, integrando las ponderaciones dado un σ^2 sobre la probabilidad de que cada valor de σ^2 es el que corresponde a la población. Suponiendo una distribución normal, la función de verosimilitud

$$f\left(y|\mathfrak{M}, \sigma^2, \mu = 0\right),$$

puede encontrarse de, al menos, dos maneras. En un método se supone que los errores de las predicciones de un conjunto de modelos son simplemente extracciones independientes de una distribución normal con media cero y varianza desconocida, donde lo que se quiere es encontrar la distribución marginal de esta varianza. El método alternativo es asumir que hay un conjunto de modelos $\mathfrak{M}$, y encontrar la distribución de probabilidad de la varianza de cada modelo y luego sumar sobre las distribuciones del conjunto de modelos para obtener la distribución marginal de la varianza. En general el resultado de estos dos métodos no será el mismo.

Si se supone que se tienen Q modelos, $\mathfrak{M} = \{M_1, ..., M_Q\}$ y, por simplicidad, se asume que cada modelo comprende de K predicciones. Entonces, con el primer método se calcula la probabilidad de un valor dado de σ^2 usando;

$$f\left(y|\mathfrak{M},\sigma^2,\mu=0\right)=\prod_{m=1}^{Q}\prod_{k=1}^{K}\left[2\pi\sigma^2\right]^{-.5}\exp\left[-.5\frac{\left(\varepsilon_{m,k}\right)^2}{\sigma^2}\right],$$

donde $\varepsilon_{m,k}$ es el error en la predicción k^{th} del modelo m^{th}. Por supuesto, dado que estamos usando una muestra finita, estas probabilidades necesitan ser normalizadas sobre σ^2 de manera que integre a uno.

El segundo método encuentra la probabilidad para un σ^2 como,

$$\begin{aligned} f\left(y|\mathfrak{M},\sigma^2,\mu=0\right) &= \sum_{m=1}^{Q} f\left(y|M_m,\sigma^2,\mu=0\right) \\ &= \sum_{m=1}^{Q}\prod_{k=1}^{K}\left[2\pi\sigma^2\right]^{-.5}\exp\left[-.5\frac{\left(\varepsilon_{m,k}\right)^2}{\sigma^2}\right], \end{aligned}$$

Donde primero se encuentra la probabilidad condicional para cada modelo y luego se suma para todos ellos. Como antes, las probabilidades necesitan ser normalizadas sobre el conjunto de σ^2 de manera que la integral sea unitaria.

La Figura 7 muestra las ponderaciones de las partes relevantes de la Figura 6, junto con la distribución de los σ^2 generados usando los dos métodos alternativos desarrollados analíticamente en esta sección (distribuciones marcadas como A y B),[8] junto con una línea vertical C que indica $\widehat{\sigma}^2$ correspondiente al método elegido Para el ejemplo analizado, los dos métodos alternativos indican que solo el modelo 10 (con una ponderacion de 1) recibiría una ponderación positiva. El método recomendado en este trabajo intercepta las líneas de ponderación al nivel en el cual se minimizan los errores de predicción agregados. Este método, en el ejemplo simulado da como resultado que cinco de los modelos reciben una ponderación significativamente diferente de cero en el promedio bayesiano.

4 Una aplicación al REM

En esta sección se aplica el método a datos reales provenientes del REM. Se supone que cada participante tiene su propio modelo independiente del cual extraer los valores esperados de las variables a pronosticar. Se utilizan las proyecciones a un mes y dos meses de la inflación mensual correspondiente a los períodos Febrero 2004 a Marzo 2006. Es necesario remarcar algunas cuestiones. Primero, dado que el REM comenzó a implementarse a comienzos del 2004, la historia de predicciones es relativamente escasa (entre 26 y 27 observaciones). Segundo, la cantidad de encuestados que alguna vez han participado en el REM es 65, pero el número que ha respondido en todos los periodos es algo menor y depende de cada variable particular (21 para las proyecciones a un mes de la inflación mensual y sólo 12 para la inflación mensual a dos meses,

[8] La altura de las dos distribuciones han sido ajustadas a la altura del gráfico, aunque las proporciones han sido mantenidas.

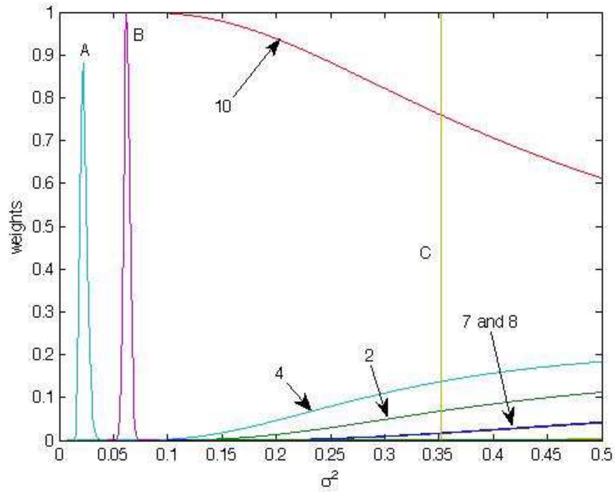

Figure 7: Distribución de los ponderadores para cada uno de los métodos propuestos para hallar σ^2.

No. datos	19	20	21	22	23	24	25	26
No. Modelos	29	27	25	24	23	17	15	12

Table 7: Número de participantes como función del número de respuestas completas (encuesta de inflación a dos meses)

ver Tabla 7). Tercero, el método tal como se ha desarrollado hasta aquí supone un conjunto de historia completa para cada firma. Muestras incompletas generan algunos inconvenientes técnicos derivados de la necesidad de comparar funciones de verosimilitud diferentes lo que complica el cálculo de las ponderaciones bayesianas. Se han realizados desarrollos en esa dirección aunque los mismos escapan al alcance del presente trabajo. Por lo tanto, en lo que sigue se muestran sólo resultados correspondientes a muestras completas, reduciéndose el conjunto de firmas participantes. [9]

La Tabla 8 muestra los resultados correspondientes a los últimas seis predicciones de inflación a dos meses empleando un conjunto completo de 12 firmas participantes. En todos los casos, salvo uno, el promedio bayesiano da la respuesta más aproximada acerca de la inflación realizada dos meses después.

Otro resultado digno de remarcar es que el promedio bayesiano demuestra ser muy sensible a la identificación de cambios en la tendencia de la variable subyacente. Esto estaría reflejando el hecho de que las firmas con mejor com-

[9] En el Apéndice 8 se muestran los resultados para la predicción a dos meses de la inflación mensual para conjuntos extendidos de participantes utilizando datos creados artificialmente. Los resultados no muestran diferencias significativas.

π_{2mth}	Valor real	Prom. Bayesiano	Mediana	Prom. Arithmetico
Oct-2005	0.8%	0.75%	0.7%	0.69%
Nov-2005	1.2%	1.0%	0.5%	0.583%
Dic-2005	1.1%	0.95%	0.95%	0.95%
Ene-2006	1.3%	1.285%	1.35%	1.275%
Feb-2006	0.4%	0.8%	1.0%	1.03%
Mar-2006	1.2%	1.208%	1.25%	1.208%

Table 8: Últimas seis proyecciones de la inflación mensual a 2 meses.

π_{1mth}	Valor real	Prom. Bayesiano	Mediana	Prom. Arithmetico
Oct-2005	0.8%	0.78%	0.8%	0.7762%
Nov-2005	1.2%	0.8%	0.7%	0.715%
Dic-2005	1.1%	1.1%	1.2%	1.15%
Ene-2006	1.3%	1.3%	1.3%	1.37%
Feb-2006	0.4%	0.8%	1.0%	0.97%
Mar-2006	1.2%	1.2%	1.2%	1.19%

Table 9: Últimas seis proyecciones de la inflación mensual a 1 mes.

portamiento histórico utilizan modelos que permiten predecir con mayor justeza cambios de tendencia. Al ponderar de manera similar todas las firmas, los cambios de tendencia no son fácilmente identificados por los demás métodos. En efecto, las proyecciones correspondientes a Noviembre 2005 y Febrero 2006 ilustran claramente este hecho. En el primer caso, las expectativas se ubicaban significativamente por debajo del valor realizado en promedio ($\sim 0.5\%$), siendo la realización (1.2%), en tanto que la predicción bayesiana arroja un valor de 1.0%, capturando el cambio de tendencia respecto a lo observado en el período anterior. En el segundo caso ocurre lo contrario, las expectativas promedio se ubican por encima del valor realizado, en tanto que el promedio bayesiano corrige parcialmente este sesgo.

La Tabla 9 muestra los resultados de los últimos seis pronósticos a un mes de la inflación mensual cuando el conjunto de información es de 21 participantes. Aquí, de nuevo, el promedio bayesiano da la mejor predicción de la inflación realizada un mes después. También permite identificar cambios de tendencia, aunque este efecto no es tan marcado, en parte porque en proyecciones a plazos más cortos los conjuntos de información disponible por los participantes tienden a converger.

5 Conclusiones

El presente trabajo muestra que el método de promedios bayesianos puede utilizarse para mejorar predicciones generadas por un conjunto de modelos. La metodología es aplicable aún cuando se desconoce el detalle de los modelos utilizados. Se utiliza el conjunto de desvíos históricos de pronósticos respecto a la variable observada, para generar los ponderadores utilizando una función de verosimilitud que se asume normal con media cero. El valor de la varianza de esta función de verosimilitud normal se elige como el que minimiza la suma de errores cuadrados de las predicciones pasadas. El método desarrollado genera resultados que son estadísticamente superiores a otros métodos usualmente utilizados para combinar pronósticos.

Adicionalmente se exploran algunas características del método, como su eficiencia relativa, la estabilidad de los ponderadores y su sensibilidad a la información contenida en los *priors* propuestos.

La aplicación del método a datos del REM del BCRA, muestra que, aún contando con relativamente poca información, el pronóstico ponderado bayesiano mejora las estimaciones de la mediana, el promedio aritmético y otros estadísticos usualmente utilizados. El método es particularmente sensible en la identificación de cambios de tendencias, ya que las firmas que han demostrado tener mejores modelos subyacentes reciben una mayor ponderación.

6 Referencias

- BCRA, (2006) *Metodología del Relevamiento de Expectativas de Mercado* (REM), February.

- Bates, J.M. and Granger, C.W.J. (1969): "The combination of forecasts". *Operational Research Quarterl*y 20, p. 451-468.

- Canova, Fabio (2006) *Methods for Applied Macroeconomic Research*, Princeton University Press, Princeton, forthcoming.

- Jacobson, Tor, and Sune Karlsson, (2004) "Finding good predictors for inflation: a Bayesian model averaging approach," *Journal of Forecasting*, John Wiley & Sons, Ltd., 23(7), p. 479-496.

- MacKay, David J. C. (2003) *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, Cambridge.

7 Apéndice A. Desempeño relativo de los métodos

A continuación se detallan los resultados para otros valores de $[\sigma_a^2, \sigma_b^2]$.

$\boldsymbol{\sigma_a^2 = 0.0, \sigma_b^2 = 0.25}$	Indice	Ranking
Prom. Bayesiano	2.90	3
Prom. Aritmético	4.59	6
Prom. Min. Varianza	2.72	1
Mejor modelo	2.83	2
"Top 5"	4.11	5
Mediana	3.84	4

$\boldsymbol{\sigma_a^2 = 0.0225, \sigma_b^2 = 0.25}$	Indice	Ranking
Prom. Bayesiano	2.94	1
Prom. Aritmético	3.96	6
Prom. Min. Varianza	3.04	2
Mejor modelo	3.68	4
"Top 5"	3.66	3
Mediana	3.72	5

$\boldsymbol{\sigma_a^2 = 0.09, \sigma_b^2 = 0.25}$	Indice	Ranking
Prom. Bayesiano	2.90	1
Prom. Aritmético	2.92	2
Prom. Min. Varianza	3.20	3
Mejor modelo	4.80	6
"Top 5"	3.74	5
Mediana	3.44	4

$\boldsymbol{\sigma_a^2 = 0.205, \sigma_b^2 = 0.25}$	Indice	Ranking
Prom. Bayesiano	2.86	2
Prom. Aritmético	2.80	1
Prom. Min. Varianza	3.22	3
Mejor modelo	4.96	6
"Top 5"	3.75	5
Mediana	3.42	4

π_{2mth} (23 Mod)	Valor real	Prom. Bayes.	Mediana	Prom. Arit.
Oct-2005	0.8%	0.764%	0.7%	0.68%
Nov-2005	1.2%	1.022%	0.7%	0.69%
Dic-2005	1.1%	0.977%	1.0%	0.78%
Ene-2006	1.3%	1.279%	1.3%	1.25%
Feb-2006	0.4%	0.8%	1.0%	1.01%
Mar-2006	1.2%	1.2%	1.2%	1.17%

Table 10: Eficiencia de los métodos como función del número de datos disponibles.

π_{2mth} (29 Mod)	Valor real	Prom. Bayes.	Mediana	Prom. Arit.
Oct-2005	0.8%	0.73%	0.7%	0.72%
Nov-2005	1.2%	1.0%	0.7%	0.74%
Dic-2005	1.1%	1.0%	1.0%	1.0%
Ene-2006	1.3%	1.23%	1.2%	1.23%
Feb-2006	0.4%	0.79%	1.0%	1.03%
Mar-2006	1.2%	1.194%	1.1%	1.19%

Table 11: Eficiencia de los métodos como función del número de datos disponibles.

8 Apéndice B. Completando artificialmente la muestra

En general, cuando el conjunto de datos no es completo existen metodologías estadísticas alternativas que pueden utilizarse para completar los mismos. No obstante, tales métodos no son del todo ajustados cuando el tamaño de la muestra disponible es considerablemente pequeño comparado con el número de observaciones en blanco. En el caso bajo análisis se opta por completar la muestra utilizando valores generados aleatoriamente a partir de una distribución normal centrada en la cantidad realizada y con varianza similar a la observada para cada participante. Debajo se muestran dos ejemplos con conjuntos de 19 y 23 participantes con datos artificialmente completados. Como puede observarse, los resultados no difieren sustancialmente a los obtenidos con la muestra reducida (pero balanceada) de 12 firmas participantes.