



“UN ENFOQUE MIDAS MODIFICADO: FB-MIDAS”

**Trabajo de Investigación presentado
para optar al Grado Académico de
Magíster en Economía**

Presentado por

Sr. Renzo Enrique Miní Cuadros

Asesor: Profesor Diego Martín Winkelried Quezada

2018

Dedicatoria

Dedicado a mi familia:

Miguel Miní, Silvia Cuadros y Diego Miní,
por su apoyo incondicional.

Agradecimiento

Agradecimiento especial a Diego Winkelried,
por su orientación y asesoramiento.

Resumen ejecutivo

Los modelos de series de tiempo tradicionales asumen una misma frecuencia entre la variable dependiente y las variables explicativas. Sin embargo, en finanzas y en macroeconomía existen variables dependientes trimestrales que pueden ser explicadas o predichas por variables independientes diarias o mensuales, respectivamente. Para resolver este problema, la literatura ha desarrollado la metodología MIDAS (*mixed-data sampling*) que emplea un polinomio de rezagos distribuidos para relacionar variables de alta frecuencia con variables de baja frecuencia. Cuando la diferencia entre frecuencias es alta, típicamente se han empleado restricciones a los coeficientes para reducir la varianza de los estimadores y solucionar el problema de sobreparametrización (metodología MIDAS). Cuando la diferencia entre frecuencias es baja, existen relativamente pocos parámetros a estimar, por lo que un modelo sin restricciones (U-MIDAS) funciona mejor para *nowcasting* y *backcasting*.

Esta investigación pretende darle un tratamiento bayesiano a la decisión de imponer o no restricciones al polinomio del modelo. Es decir, trata de ubicarse en el medio de los dos extremos (MIDAS y U-MIDAS) al imponer restricciones, determinadas empíricamente, de manera estocástica. El tratamiento consiste de un *prior* de “suavizamiento” sobre la distribución de rezagos del polinomio, y para encontrarlo, realiza tanto una calibración bayesiana empírica como un promedio ponderado de modelos bayesianos.

Índice

Índice de gráficos.....	vii
Índice de anexos.....	viii
Capítulo I. Introducción.....	1
1. Problema-motivación.....	1
2. Soluciones <i>a priori</i>	1
3. MIDAS y U-MIDAS.....	2
4. Propuesta FB-MIDAS, objetivo e hipótesis.....	3
Capítulo II. Revisión de la literatura.....	4
1. Polinomios de rezagos distribuidos (PRD).....	4
2. Metodología MIDAS.....	5
3. Metodología U-MIDAS.....	6
4. Propuesta FB-MIDAS.....	7
5. Tratamiento bayesiano.....	12
6. Propiedades en muestras pequeñas.....	14
7. Error cuadrático medio de predicción (ECMP).....	15
Capítulo III. Metodología.....	17
1. Elección de modelos.....	17
2. Especificación de modelos.....	19
3. Comparación de modelos.....	21
Capítulo IV. Ejercicio de evaluación.....	23
1. Procesos generadores de datos (PGD).....	23
2. Data empírica.....	24
3. Estrategia empírica.....	24
Capítulo V. Análisis de resultados.....	26
1. Disparidad trimestral-mensual.....	26
2. Disparidad trimestral-diaria.....	28
3. Disparidad mensual-diaria.....	30

Conclusiones y recomendaciones	34
1. Conclusiones.....	34
2. Recomendaciones	34
Bibliografía.....	35
Anexos.....	37
Nota biográfica.....	45

Índice de gráficos

Gráfico 1.	VAR (1).....	26
Gráfico 2.	PBI real trimestral vs. indicadores líderes mensuales	27
Gráfico 3.	MIDAS PGD	28
Gráfico 4.	Consumo real trimestral vs. indicadores líderes diarios	29
Gráfico 5.	U-MIDAS PGD	30
Gráfico 6.	PBI real mensual vs. indicadores líderes diarios	31

Índice de anexos

Anexo 1.	Error cuadrático medio de predicción (ECMP)	38
Anexo 2.	Verosimilitud marginal de la data	39
Anexo 3.	Predicciones de data empírica con la metodología FB-MIDAS bajo diferentes disparidades	40
Anexo 4.	Valores de lambda óptimo por BMS	43

Capítulo I. Introducción

1. Problema-motivación

Las series de tiempo económicas difieren sustancialmente por la frecuencia con la que se presentan. Por ejemplo, las variables financieras son observadas, regularmente, de manera diaria o intradiaria; por otro lado, algunas variables macroeconómicas presentan frecuencia mensual (producción manufacturera, inflación, agregados monetarios, etc.), mientras que los agregados reales de las cuentas nacionales (PBI, consumo, inversión, etc.) están disponibles solo de manera trimestral (Ghysels, Sinko y Valkanov 2007). Adicionalmente, sucede que debido a rezagos en publicación y revisiones, las realizaciones de algunas variables macroeconómicas en períodos dados pueden estar ausentes. Por ejemplo, en el caso peruano, la data de PBI trimestral está disponible públicamente 45 días después del término del trimestre respectivo, mientras que en EE.UU. y la Eurozona, los rezagos pueden durar hasta 2 trimestres (Winkelried 2012).

Estos problemas de frecuencia mixta de muestreo y de rezagos en publicación constituyen un desafío para los hacedores de política cuando realizan análisis empíricos de relaciones entre variables económicas para tomar decisiones: cómo usar la data disponible de la mejor manera posible para predecir fuera de la muestra, en tiempo real, la variable de baja frecuencia¹. Dicho desafío parte, principalmente, de que el análisis econométrico tradicional de series de tiempo asume una misma frecuencia entre la variable dependiente y sus explicativas.

2. Soluciones *a priori*

La solución más simple es trabajar en la frecuencia más baja de la data. Esto requiere una agregación o interpolación de variables de alta frecuencia con una potencial pérdida de información relevante o con una imposición de supuestos usualmente poco plausibles, que pueden dificultar el análisis de la relación entre las variables (Forni, Marcellino y Schumacher 2015). Como una alternativa, se puede formular un modelo de espacio de estados, que trata como variables latentes a los valores subyacentes de alta frecuencia de las variables de baja frecuencia, y usar el filtro de Kalman para encontrar la data faltante (Forni y Marcellino 2013).

Para tomar en cuenta la naturaleza de frecuencia mixta de la data, las variables de baja frecuencia son interpoladas de acuerdo con su naturaleza de flujo o de *stock* (Proietti 2006). Si bien este procedimiento permite realizar predicciones de manera sencilla, requiere especificar por completo

¹ “Predecir”, en este contexto, significa 2 cosas: *nowcasting* (en un período particular de la variable de alta frecuencia, predecir la variable de baja frecuencia para el período corriente) y *backcasting* (predecir el dato, aún no disponible, de la variable de baja frecuencia del período anterior).

una estructura dinámica lineal que describa la evolución de la variable latente, además de precisar cómo se agrega esta (flujo o *stock*) (Bai, Ghysels y Wright 2013).

Como otra alternativa, sin perder información de alta frecuencia, y que también se adecúa fácilmente para ejercicios de predicción, se encuentran las regresiones de series de tiempo *mixed data-sampling* (MIDAS) (Armesto, Engemann y Owyang 2010). Se basan, típicamente, en polinomios de rezagos distribuidos (PRD). A diferencia de los espacios de estados, MIDAS predice directamente y utiliza métodos estándares de regresión al formular una sola ecuación en forma reducida. Además, no requiere modelar el comportamiento de los indicadores de alta frecuencia (solo cómo se vinculan en el agregado de baja frecuencia).

3. MIDAS y U-MIDAS

Las regresiones MIDAS han sido aplicadas en la literatura financiera en el contexto de predicciones de volatilidad (Ghysels, Santa-Clara y Valkanov 2004). En la literatura macroeconómica, las aplicaciones se han utilizado normalmente para *nowcasting* y *forecasting* (Clements y Galvão 2009): predecir el crecimiento trimestral del PBI usando indicadores mensuales del ciclo económico (Clements y Galvão 2008). Finalmente, existe literatura que utiliza variables financieras diarias para predecir el PBI trimestral (Andreou, Ghysels y Kourtellos 2010).

Estas regresiones han utilizado el rezago exponencial de Almon y, debido a la no-linealidad del PRD, han sido estimadas mediante mínimos cuadrados no lineales (MCNL). Dicha metodología (MIDAS)² impone restricciones particulares (*priors* de “suavizamiento” determinísticos) para asegurar especificaciones parsimoniosas, y evitar sobreparametrizaciones, cuando la disparidad entre frecuencias de las variables dependiente e independientes es grande (por ejemplo, trimestral-diaria, como suele suceder con data financiera) (Kuzin, Marcellino Schumacher 2011). Sin embargo, cuando la disparidad es pequeña (por ejemplo, trimestral-mensual, como normalmente ocurre con data macroeconómica), se han utilizado PRD lineales no restringidos (dejando de lado PRD funcionales) que no requieren MCNL y pueden ser estimados por mínimos cuadrados ordinarios (MCO) (Forni, Marcellino y Schumacher 2015). Esta última metodología (U-MIDAS)³ no necesita imponer restricciones funcionales ya que el modelo no suele estar sobreparametrizado. Por un lado, MIDAS introduce sesgo de variable omitida al restringir estimadores y subestima la varianza muestral de los restantes, mientras que por otro

² Que representa: *Restricted* MIDAS.

³ Que representa: *Unrestricted* MIDAS.

lado, U-MIDAS no introduce sesgo pero sobreestima la varianza muestral, con lo cual la precisión de los estimadores es pobre.

4. Propuesta FB-MIDAS, objetivo e hipótesis

El propósito de la investigación es evaluar la *performance* relativa de una variante (FB-MIDAS)⁴ vs. las metodologías clásicas MIDAS y U-MIDAS ya conocidas. Esta radica en una estimación mixta de Theil y Goldberger (1961), que plantea restricciones estocásticas para complementar la información limitada de una muestra pequeña, con el objetivo de economizar los grados de libertad. Es decir, esta propuesta se ubica en un intermedio entre las anteriores⁵: MIDAS impone restricciones determinísticas (tal vez muy fuertes), U-MIDAS no impone restricciones (de manera también determinística), mientras que FB-MIDAS impondría *priors* de “suavizamiento” estocásticos implementados a través de criterios bayesianos. Cabe resaltar que la imposición de dichos *priors* estocásticos, al asumir restricciones más débiles, pero asegurando modelos no sobreparametrizados, genera típicamente estimadores sesgados (menos que MIDAS), pero con menores varianzas que U-MIDAS, lo cual abre la primera hipótesis de obtener ganancias en términos de error cuadrático medio sobre estas dos metodologías. Finalmente, dada la primera hipótesis, se podría establecer una segunda respecto de un *ranking* de ECM: para disparidad trimestral-mensual entre la variable dependiente y sus explicativas, FB-MIDAS seguida de U-MIDAS y luego MIDAS; mientras que, para disparidad trimestral-diaria, FB-MIDAS seguida de MIDAS y luego U-MIDAS.; y, para disparidad mensual-diaria el *ranking* dependería del PGD.

En el capítulo II, se especifica el modelo MIDAS autorregresivo de rezagos distribuidos (MARD); se presentan los enfoques MIDAS y U-MIDAS clásicos, y FB-MIDAS propuesto; se expone un tratamiento bayesiano para ellos; se analizan las propiedades de los tres estimadores en muestras pequeñas; y se revisa el *trade-off* sesgo-varianza en el error cuadrático medio de predicción (ECMP). En el capítulo III, se plantean metodologías de elección de modelos, especificación de modelos y comparación de modelos. En el capítulo IV, se diseña un ejercicio de evaluación robusto que involucra simulaciones de Montecarlo con 3 diferentes procesos generadores de datos (PGD), además de data empírica macroeconómica. En el quinto capítulo se presentan y analizan los resultados para las tres disparidades de frecuencias mencionadas. Finalmente, se cierra con las conclusiones de la investigación y se proponen recomendaciones o extensiones plausibles a futuro.

⁴ Que representa: *Flexible Bayesian* MIDAS.

⁵ Winkelried (2012) desarrolla la misma idea, pero sin presentar un enfoque ni metodología bayesianos.

Capítulo II. Revisión de la literatura

1. Polinomios de rezagos distribuidos (PRD)

En un modelo MIDAS típico, cada período de baja frecuencia está compuesto por un número arbitrario m de alta frecuencia. Para mantener la notación simple, se utiliza el índice t en la frecuencia más alta y se maneja la data de baja frecuencia como observable solo cuando t es múltiplo de m , y no disponible de otro modo.

Sea y_t una variable observable solo cada m períodos (esto es, para una muestra de tamaño n de la variable de baja frecuencia, y_t es observable solo para $t = m, 2m, 3m, \dots, nm$). Por otro lado, sea x_t un indicador de alta frecuencia, observable para $t = 1, 2, 3, \dots, nm$. Así, el enfoque MIDAS consiste en formular un modelo de rezagos distribuidos (MRD) que relaciona y_t con x_t y sus rezagos⁶:

$$\begin{aligned} y_t &= \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots + \beta_q x_{t-q} + \varepsilon_t && \text{para } t \\ &= m, 2m, 3m, \dots, nm && (1) \end{aligned}$$

donde ε_t es un término de error no autocorrelacionado y homocedástico. El parámetro $q \geq 0$ es el tamaño del rezago en la relación lineal dinámica entre y_t y x_t . Este modelo puede ser escrito en forma matricial como $y = X\beta + \varepsilon$, donde y y ε son vectores de dimensión $n \times 1$ y X es una matriz de regresores de dimensión $n \times (q + 1)$.

Este enfoque es fácilmente adaptable a diferentes estructuras de data con el objetivo de formar una predicción con toda la información disponible en un período dado. Sea el interés de la investigación el predecir el valor de y_t cuando su última observación disponible es $y_{t-(\kappa+1)m}$. El parámetro κ , el rezago en publicación, controla el período entre el último dato disponible y el dato a predecir: si $\kappa = 0$, entonces la última observación de la variable dependiente corresponde al período anterior y_{t-m} , mientras que si $\kappa = 1$, la última observación corresponde a dos períodos anteriores y_{t-2m} , y así sucesivamente. Adicionalmente, considerar que el último dato disponible del indicador de alta frecuencia es x_{t+h} . El parámetro h determina la naturaleza de la predicción: si $-m < h \leq 0$, entonces x_{t+h} y sus rezagos contienen información dentro del período de baja frecuencia t (antes de que termine), con lo cual la predicción significa *nowcasting*; si $h > 0$, entonces x_{t+h} y sus rezagos contienen información más allá del período de baja

⁶ La presente especificación ha sido extraída en su totalidad de Winkelried (2012).

frecuencia t (después de haber terminado), con lo cual la predicción significa *backcasting*.
 Con estas consideraciones, el MRD básico en (1) puede ser extendido a un MARD de la forma:

$$\begin{aligned}
 & y_t \\
 & = \beta_0 x_{t+h} + \beta_1 x_{t+h-1} + \beta_2 x_{t+h-2} + \dots + \beta_q x_{t+h-q} + \gamma_1 y_{t-(\kappa+1)m} + \gamma_2 y_{t-(\kappa+2)m} + \dots \\
 & + \gamma_p y_{t-(\kappa+p)m} \\
 & + \varepsilon_t
 \end{aligned}$$

donde $p \geq 0$ es el número de términos autorregresivos (de la variable de baja frecuencia)⁷.

Cabe precisar que el propósito de este MARD en (2) es predecir y_t usando la información disponible ($y_{t-(\kappa+1)m}$, x_{t+h} y los rezagos de x_{t+h}). Así, dada la configuración ya presentada, puede suceder que el dato x_s esté disponible en algún momento del tiempo pero el dato y_s no lo esté. O también puede suceder que y_t esté sujeta a rezagos en publicación y revisiones mientras que x_t no lo esté. Sin embargo, en cualquier caso la predicción es directa: si b es un vector de estimadores de β , esta predicción está dada por $\hat{y}_s = x_s b$.

2. Metodología MIDAS

Un problema principal con (2), si no se restringe, es la sobreparametrización: ya que los regresores son observados en una frecuencia más alta que la variable dependiente, se pueden llegar a requerir varios rezagos (q alto relativo a n) para modelar adecuadamente la relación dinámica entre y_t y x_t . Por ejemplo, si y_t es trimestral y está afectada por tres trimestres de rezagos de una variable x_t diaria, entonces se necesitan $q = 3 \times 3 \times 30 = 270$ rezagos en la variable de alta frecuencia (Ghysels, Sinko y Valkanov 2007).

El enfoque MIDAS clásico busca solucionar este problema al especificar un PRD funcional que implica un modelo que permite largos rezagos parsimoniosamente. Sea este:

$$\begin{aligned}
 \gamma(L^m)\omega(L)y_t & = \beta_1 B(L; \theta)x_t + \varepsilon_t && \text{para } t \\
 & = m, 2m, 3m, \dots, nm && (3)
 \end{aligned}$$

donde $\gamma(L^m) = 1 - \gamma_1 L^m - \dots - \gamma_p L^{(\kappa+p)m}$ es un polonimo en el operador de rezagos con baja frecuencia, $\omega(L) = \omega_0 + \omega_1 L + \dots + \omega_{m-1} L^{m-1}$ es un operador que caracteriza el esquema de

⁷ Se pueden imponer, adicionalmente, restricciones de factores comunes en la estimación de los coeficientes γ para evitar discontinuidades (cada m períodos) en la función de impulso-respuesta de x_t en y_t (Clements y Galvão 2008).

agregación temporal de la variable dependiente⁸, $B(L; \theta) = \sum_{k=0}^K b(k; \theta)L^k$ y L es el operador de rezagos de alta frecuencia tal que $L^s x_t = x_{t-s}$. Así, los rezagos son función de un vector de parámetros θ a ser estimados, y β_1 es identificado exigiendo que $B(L; \theta) = 1$.

$b(k; \theta)$ se puede parametrizar de diferentes maneras. El enfoque clásico, en particular, ha utilizado el polinomio finito del rezago exponencial de Almon (Clements y Galvão 2008), el cual exige:

$$\begin{aligned}
 b(k; \theta) &= \frac{e^{\theta_0 + \theta_1 k + \theta_2 k^2 + \dots + \theta_d k^d}}{\sum_{k=1}^K e^{\theta_0 + \theta_1 k + \theta_2 k^2 + \dots + \theta_d k^d}} = \alpha_0 + \alpha_1 k + \alpha_2 k^2 + \dots + \alpha_d k^d \\
 &= \beta_k \quad \text{para } k \\
 &= 0, 1, 2, \dots, q
 \end{aligned} \tag{4}$$

donde si $q > d$, existirán menos parámetros α que β , lo que significa que (4) impone $q - d$ restricciones determinísticas en β .

Es decir, dicho enfoque especifica que los coeficientes β_k sean funciones no lineales suavizadas de k^9 , lo cual introduce no linealidades en el espacio de parámetros, por lo que una versión restringida de (2) puede ser estimada por MCNL.

3. Metodología U-MIDAS

Como se ha mencionado, la sobreparametrización es la mayor limitación cuando la disparidad entre frecuencias es grande (común en aplicaciones financieras). Sin embargo, cuando la disparidad entre frecuencias es pequeña (por ejemplo, trimestral-mensual, común en aplicaciones macroeconómicas) y la muestra es grande, el enfoque MIDAS no lineal puede ser muy restrictivo.

Por ejemplo, si y_t es trimestral y está afectada por tres trimestres de rezagos de una variable x_t mensual, entonces se necesitan $q = 3 \times 3 = 9$ rezagos en la variable de alta frecuencia. Por otro lado, puede suceder simplemente que la restricción funcional $\beta_1 B(L; \theta)$ no es válida y que la aproximación con el rezago de Almon no es lo suficientemente general. Adicionalmente, si la función de impulso-respuesta de x_t a y_t es de corta duración y solo unos pocos rezagos de alta frecuencia son necesarios para capturar la dinámica, un modelo sin restricciones como el siguiente, que incluye al restringido y que puede ser estimado por MCO, es suficiente (Forni,

⁸ $\omega(L) = 1 + L + \dots + L^{m-1}$ para variables que representan un flujo y $\omega(L) = 1$ para variables que representan un *stock*.

⁹ Impone un *prior* de “suavizamiento” determinístico.

Marcellino y Schumacher 2015)¹⁰:

$$\begin{aligned} \gamma(L^m)\omega(L)y_t &= \beta(L)x_t + \varepsilon_t && \text{para } t \\ &= m, 2m, 3m, \dots, nm && (5) \end{aligned}$$

donde $\beta(L) = \beta_0 + \beta_1L + \dots + \beta_qL^q$ es un polinomio típico en el operador de rezagos.

4. Propuesta FB-MIDAS

A pesar de lo anterior, cuando la muestra es pequeña relativo al número de rezagos (independientemente a si la disparidad entre frecuencias es grande o pequeña), restringir la estimación puede ser deseable para economizar los grados de libertad y atenuar altas varianzas de los estimadores. Sin embargo, debido a la posibilidad de introducir sesgo de variable omitida por imponer restricciones determinísticas que pueden ser muy fuertes, tomar una ruta intermedia e imponer restricciones estocásticas puede ser una alternativa. El asumir restricciones más débiles (que MIDAS) pero asegurando modelos no tan sobreparametrizados (como U-MIDAS) genera, típicamente, estimadores menos sesgados que MIDAS pero con menores varianzas que U-MIDAS, lo cual abre la posibilidad de obtener ganancias en términos de error cuadrático medio tanto sobre MIDAS como sobre U-MIDAS para cualquier disparidad entre frecuencias.

Este enfoque intermedio se basa en la noción de *priors* de “suavizamiento” estocásticos (Shiller 1973)¹¹, en donde cada coeficiente en el MRD está restringido a satisfacer aproximadamente el rezago exponencial de Almon de grado d :

$$\begin{aligned} \beta_k &= \alpha_0 + \alpha_1k + \alpha_2k^2 + \dots + \alpha_dk^d + \omega_k && \text{para } k \\ &= 0, 1, 2, \dots, q && (6) \end{aligned}$$

donde ω_k es un término estocástico con media cero¹² y $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_d$ son coeficientes a ser estimados así como en (4). Notar que estas restricciones, a diferencia de (4), son estocásticas (no necesariamente se cumplirán), lo cual es ventajoso ya que d puede ser pequeño sin imponer restricciones muy fuertes a la data.

Una manera conveniente de obtener un estimador b sujeto a (6) es mediante la técnica mixta de

¹⁰ Equivalente a (2) cuando $\omega(L) = 1$.

¹¹ En palabras de Shiller, los *priors* de “suavizamiento” determinísticos son especificados no porque uno confíe en la exactitud de ellos, sino que confía en la suavidad de la distribución de rezagos. Esto da pie a introducir *priors* de “suavizamiento” estocásticos (Shiller 1973).

¹² Lo cual asegura una restricción y un estimador “mixto” (que se observará a continuación) insesgados.

Theil y Goldberger (1961) (frecuentista con espíritu bayesiano).

Considere el siguiente modelo general que utiliza un MRD finito con coeficientes estocásticos:

$$\begin{aligned}
 y_t &= \sum_{k=0}^q \beta_{kt} x_{t-k} + \varepsilon_t && \text{para } t \\
 &= m, \dots, nm && (7)
 \end{aligned}$$

donde ε_t es el término usual de error correspondiente a la t-ésima observación. Se puede notar que β_{kt} es un coeficiente que varía estocásticamente tal que:

$$\begin{aligned}
 &\beta_{kt} \\
 &= \beta_k \\
 &+ \varepsilon_{kt}
 \end{aligned}$$

donde ε_{kt} representa una distorsión no observable y β_k se comporta como en (6). De esta manera, se observa que si ε_{kt} y ω_k son constantes (pueden ser cero) para todo k y todo t , entonces (8) se reduce a una especificación de Almon como la expuesta para el caso MIDAS.

Se realizan los siguientes supuestos¹³:

Supuesto 1:

- (i) $E(\varepsilon_t) = E(\varepsilon_{kt}) = E(\omega_k) = 0 \quad \forall k, t$
- (ii) $E(\varepsilon_t \varepsilon_{kt}) = E(\varepsilon_t \omega_k) = E(\omega_k \varepsilon_{kt}) = 0 \quad \forall k, t$
- (iii) $E(\omega_k \omega_i) = \sigma_\omega^2 \quad \text{si } k = i$
 $= 0 \quad \text{d. o. m}$
- (iv) $E(\varepsilon_t \varepsilon_{t'}) = \sigma_\varepsilon^2 \quad \text{si } t = t'$
 $= 0 \quad \text{d. o. m}$
- (v) $E(\varepsilon_{kt} \varepsilon_{k't'}) = \sigma_{kk}^2 \quad \text{si } t = t' \wedge k = k'$
 $= 0 \text{ d.o.m.}^{14}$
- (vi) x_{t-k} es una variable independiente distribuida independientemente de ε_t , ε_{kt} y ω_k para todo k y todo t .

¹³ El siguiente desarrollo ha sido extraído, por completo, de Ullah y Raj (1979) y de Ullah y Raj (1980).

¹⁴ Se puede extender esta especificación para el caso en el que los ε 's siguen un proceso ARMA.

Si se sustituye (8) en (7), se obtiene:

$$y_t = \sum_{k=0}^q \beta_k x_{t-k} + \varepsilon_t \quad \text{para } t$$

$$= m, \dots, nm \quad (9)$$

donde:

$$\varepsilon_t$$

$$= \varepsilon_t$$

$$+ \sum_{k=0}^q \epsilon_{kt} x_{t-k}$$

tal que, tras el Supuesto 1:

(a) $E(\varepsilon_t) = 0 \quad \forall t$

(b) $E(\varepsilon_t \varepsilon_{t'}) = \Omega_{tt} \quad \text{si } t = t'$

$$= 0 \quad \text{d.o.m}$$

y además, Ω_{tt} tiene la siguiente forma:

$$\Omega_{tt} = \sigma_\varepsilon^2 + \sum_{k=0}^q x_{t-k}^2 \sigma_{kk}^2$$

$$> 0$$

El modelo anterior (7) y su restricción (8) pueden ser escritos, respectivamente, en forma matricial de la siguiente manera:

$$Y$$

$$= X\beta$$

$$+ \varepsilon$$

$$\beta$$

$$= A\alpha$$

$$+ \omega$$

donde:

$$\begin{aligned}
Y_{(nm-m+1) \times 1} &= \begin{bmatrix} y_m \\ \vdots \\ y_{nm} \end{bmatrix}, \beta_{(q+1) \times 1} = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_q \end{bmatrix}, \varepsilon_{(nm-m+1) \times 1} = \begin{bmatrix} \varepsilon_m \\ \vdots \\ \varepsilon_{nm} \end{bmatrix}, \\
\alpha_{(d+1) \times 1} &= \begin{bmatrix} \alpha_0 \\ \vdots \\ \alpha_d \end{bmatrix}, \omega_{(q+1) \times 1} = \begin{bmatrix} \omega_0 \\ \vdots \\ \omega_q \end{bmatrix} \\
X_{(nm-m+1) \times (q+1)} &= \begin{bmatrix} x_{m0} & \cdots & x_{mq} \\ \vdots & \ddots & \vdots \\ x_{nm0} & \cdots & x_{nmq} \end{bmatrix}, A_{(q+1) \times (d+1)} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & 1 & \ddots & \vdots \\ 1 & q & \cdots & q^d \end{bmatrix}
\end{aligned}$$

y, además, A representa la matriz de transformación de Almon con rango columna completo¹⁵: $A_{ij} = (i-1)^{j-1} \quad \forall \quad i = 1, 2, \dots, q+1$ y $j = 1, 2, \dots, d+1$; y ω representa un vector de errores con media cero y homocedásticos ($E(\omega) = 0$; $var(\omega) = \sigma_\omega^2 I_{q+1}$). Por otro lado, a partir de (11), se tiene que:

$$\begin{aligned}
E(\varepsilon) &= 0 \\
E(\varepsilon\varepsilon') &= \begin{bmatrix} \Omega_{mm} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Omega_{nmnm} \end{bmatrix} = \Omega_{(nm-q) \times (nm-q)}
\end{aligned}$$

donde Ω es una matriz definida positiva. Se puede notar que si $\sigma_{00} = \cdots = \sigma_{qq} = 0$ y $\sigma_\omega^2 = 0$, entonces el modelo es un PRD funcional con especificación de Almon (como en (3), coeficientes fijos en el tiempo y *priors* determinísticos). En este caso, la matriz de covarianzas de ε se reduce a $\sigma_\varepsilon^2 I$ y (13) se convierte en $\beta = A\alpha$ ¹⁶.

Si lo último no se asume, (13) se puede reexpresar de la siguiente manera:

$$\begin{aligned}
R\beta \\
= R\omega
\end{aligned}$$

donde R es una matriz de diferencias $(d+1)$, de rango fila completo y de dimensiones $(q-d) \times (q+1)$, definida como:

¹⁵ Se asume que los rangos de las matrices X y A son $(q+1) < (nm-m+1)$ y $(d+1) < (q+1)$, respectivamente. En particular, si $d < q$, entonces el rango de A es $(d+1)$.

¹⁶ O, como se observará inmediatamente, $R\beta = 0$.

$$R_{(q-d) \times (q+1)} = \begin{bmatrix} C_1 & C_2 & C_3 & \cdots & C_{d+2} & 0 & 0 & \cdots & 0 \\ 0 & C_1 & C_2 & \cdots & C_{d+1} & C_{d+2} & 0 & \cdots & 0 \\ 0 & 0 & C_1 & \cdots & C_d & C_{d+1} & C_{d+2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & C_1 & C_2 & C_3 & \cdots & C_{d+2} \end{bmatrix},$$

$$C_i = (-1)^{i-1} \binom{d+1}{i-1}$$

Por lo que se obtiene:

$$RA\alpha = 0$$

Se puede notar, además, que $R\omega$ es un vector aleatorio con (usando el Supuesto 1) $E(R\omega) = 0$ y $var(R\omega) = E(R\omega\omega'R') = \sigma_\omega^2 RR'$ ¹⁷.

La estimación de β se basa en un modelo concatenado que combina información muestral (12) con información estocástica *a priori* (la restricción (13)) (Rao, Toutenburg, Shalabh y Heumann 2008):

$$\begin{bmatrix} Y \\ 0 \end{bmatrix} = \begin{bmatrix} X \\ R \end{bmatrix} \beta + \begin{bmatrix} \varepsilon \\ R\omega \end{bmatrix} \text{ con } \begin{bmatrix} \varepsilon \\ R\omega \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}; \begin{bmatrix} \Omega & 0 \\ 0 & \mu RR' \end{bmatrix} \right)$$

donde $\mu = 1/\sigma_\omega^2$. Dicho estimador mixto es equivalente al estimador de mínimos cuadrados generalizados (MCG) de β en (16), que se llamará estimador Shiller Generalizado Aleatorio (SGA):

$$\begin{aligned} b_{SGA} &= (X'\Omega^{-1}X + \mu R'(RR')^{-1}R)^{-1} X'\Omega^{-1}Y \\ &= b_{MCG} + (X'\Omega^{-1}X)^{-1} R'(\mu RR' + R(X'\Omega^{-1}X)^{-1}R')^{-1} (0 - Rb_{MCG}) \end{aligned} \quad (17)$$

¹⁷ La especificación de Shiller es menos general que (19), ya que asume $R\omega$ como un vector aleatorio con media cero y matriz de covarianzas $\sigma_\omega^2 I$. En realidad, comienza con $\beta = A\alpha$, lo cual implica $R\beta = 0$, y luego impone un vector estocástico, ω , para especificar $R\beta = \omega$. Adicionalmente, en su caso $\Omega = \sigma_\varepsilon^2 I$.

Con algebra simple pero tediosa, si se asume que $\sigma_\omega^2 \rightarrow 0$, el estimador SGA puede convertirse en el estimador Almon Aleatorio (AA)¹⁸:

$$\begin{aligned} b_{AA} &= A((XA)'\Omega^{-1}(XA))^{-1}(XA)'\Omega^{-1}Y \\ &= b_{MCG} + (X'\Omega^{-1}X)^{-1}S'(S(X'\Omega^{-1}X)^{-1}S')^{-1}(0 \\ &\quad - Sb_{MCG}) \end{aligned} \quad (18)$$

Adicionalmente, si se asumen coeficientes fijos en el tiempo ($\sigma_{kk} = 0 \quad \forall k = 0, \dots, q$), lo que implica que $\Omega = \sigma_\varepsilon^2 I$, entonces el estimador AA se convierte en el estimador Almon (A):

$$\begin{aligned} b_A &= A((XA)'(XA))^{-1}(XA)'Y \\ &= b_{MCO} + (X'X)^{-1}S'(S(X'X)^{-1}S')^{-1}(0 \\ &\quad - Sb_{MCO}) \end{aligned} \quad (19)$$

Por otro lado, si se asume que $var(R\omega) = \sigma_\omega^2 I_{q+1}$, el estimador SGA puede convertirse en el estimador Shiller Aleatorio (SA):

$$\begin{aligned} b_{SA} &= (X'\Omega^{-1}X + \mu R'R)^{-1}X'\Omega^{-1}Y \\ &= b_{MCG} + (X'\Omega^{-1}X)^{-1}R'(\mu I + R(X'\Omega^{-1}X)^{-1}R')^{-1}(0 \\ &\quad - Rb_{MCG}) \end{aligned} \quad (20)$$

Y este último, bajo el supuesto de coeficientes fijos en el tiempo ($\sigma_{kk} = 0 \quad \forall k = 0, \dots, q$), lo que implica que $\Omega = \sigma_\varepsilon^2 I$, se convierte en el estimador Shiller (S):

$$\begin{aligned} b_S &= (X'X + \lambda R'R)^{-1}X'Y \\ &= b_{MCO} + (X'X)^{-1}R'(\mu I + R(X'X)^{-1}R')^{-1}(0 \\ &\quad - Rb_{MCO}) \end{aligned} \quad (21)$$

donde $\lambda = \sigma_\varepsilon^2 \mu = \sigma_\varepsilon^2 / \sigma_\omega^2$ es un parámetro que puede ser interpretado como la cantidad de información que introducen las restricciones estocásticas (en pocas palabras, el parámetro de

¹⁸ Esto sucede al notar que $R'(RR')^{-1}R = S = I - A'(A'A)^{-1}A'$, donde S es una matriz idempotente; y al usar el resultado $(X'\Omega^{-1}X + \mu S)^{-1} = (X'\Omega^{-1}X + \mu S^2)^{-1} = (X'\Omega^{-1}X)^{-1}(I - S(S'(X'\Omega^{-1}X)^{-1}S + \frac{1}{\mu}I)^{-1}S'(X'\Omega^{-1}X)^{-1})$. Dado que $b_{SGA} = (X'\Omega^{-1}X + \mu S)^{-1}X'\Omega^{-1}Y$, entonces $b_{SGA} = (X'\Omega^{-1}X + \mu S^2)^{-1}X'\Omega^{-1}Y = (X'\Omega^{-1}X)^{-1}(I - S(S'(X'\Omega^{-1}X)^{-1}S + \frac{1}{\mu}I)^{-1}S'(X'\Omega^{-1}X)^{-1})X'\Omega^{-1}Y$. Si $\sigma_\omega^2 \rightarrow 0$, entonces se obtiene el estimador b_{AA} (cuya expresión en (24) es indiferente a si emplea S o R).

“suavizamiento”).

La investigación utilizará la versión no aleatoria del estimador SGA¹⁹, es decir, el estimador Shiller Generalizado (SG):

$$\begin{aligned}
 b_{SG} &= (X'X + \lambda R'(RR')^{-1}R)^{-1}X'Y \\
 &= b_{MCO} + (X'X)^{-1}R'(\mu RR' + R(X'X)^{-1}R')^{-1}(0 \\
 &\quad - Rb_{MCO})
 \end{aligned} \tag{22}$$

Dados R y λ , el estimador SG es fácil de ser computado. Notar, además, que es lo suficientemente general para los objetivos de la investigación: si $\sigma_{\omega}^2 \rightarrow 0$, como se ha mencionado previamente, las restricciones (6), (13) y (14) son vinculantes, $\lambda \rightarrow \infty$ y $b_{SG} \rightarrow b_A$ (el estimador A o estimador MCNL del modelo restringido (3)), es decir, el estimador de la metodología MIDAS (Ullah & Raj, 1980). Por el contrario, si $\sigma_{\omega}^2 \rightarrow \infty$, las mismas restricciones son redundantes, ya que requerirían que β_k tome cualquier valor real. Como consecuencia, $\lambda \rightarrow 0$ y $b_{SG} \rightarrow b_{MCO}$ (el estimador MCO del modelo no restringido (5)), es decir, el estimador de la metodología U-MIDAS.

5. Tratamiento bayesiano

El estimador SG (de la metodología FB-MIDAS propuesta), así como el estimador A (de la metodología MIDAS) y el estimador MCO (de la metodología U-MIDAS), pueden también ser encontrados de manera bayesiana (Giles 1975).

Para el cualquier caso, se puede asumir una verosimilitud con distribución Normal como:

$$Y \sim \mathcal{N}(X\beta, \sigma_{\varepsilon}^2 I)$$

tal que:

$$\begin{aligned}
 p(Y|\beta, p, q, \sigma_{\varepsilon}^2) \\
 &= \left(\frac{1}{2\pi\sigma_{\varepsilon}^2} \right)^{\frac{nm}{2}} e^{-\frac{1}{2\sigma_{\varepsilon}^2} \|Y - X\beta\|^2}
 \end{aligned}$$

En particular, para el primer caso (estimador SG, FB-MIDAS) se puede asumir una distribución *a priori* (conjugada) de β de la siguiente manera:

¹⁹ Con lo cual, no es necesario utilizar un estimador MCO (insesgado, pero ineficiente) o MCG (insesgado y eficiente) de σ_{ε}^2 .

$$\beta \sim \mathcal{N}(A\alpha, \sigma_\beta^2 I = \sigma_\omega^2 I)$$

así como un *prior* de Jeffrey para σ_ε^2 :

$$p(\sigma_\varepsilon^2) \propto \frac{1}{\sigma_\varepsilon^2}$$

tal que:

$$p(\beta, \sigma_\varepsilon^2 | \lambda, p, q, d) = \left(\frac{1}{2\pi\sigma_\varepsilon^2} \right)^{\frac{p+q+1}{2}} |\lambda R'(RR')^{-1}R|^{\frac{1}{2}} e^{-\frac{\lambda}{2\sigma_\varepsilon^2} \beta'R'(RR')^{-1}R\beta} \left(\frac{1}{\sigma_\varepsilon^2} \right)$$

Se puede encontrar, usando el Teorema de Bayes, la distribución *a posteriori* de β con el siguiente teorema:

Teorema 1 (Maddala 1974):

Si $Y \sim \mathcal{N}(X\beta, \sigma_\varepsilon^2 I)$ y $\beta \sim \mathcal{N}(A\alpha, \sigma_\beta^2 I = \sigma_\omega^2 I)$, entonces la distribución *a posteriori* de β ($p(\beta|Y, \sigma_\varepsilon^2)$) es bien comportada y se define como $\mathcal{N}((X'X + \lambda R'(RR')^{-1}R)^{-1}X'Y, \sigma_\varepsilon^2(X'X + \lambda R'(RR')^{-1}R)^{-1})$.

De esta manera, la media (o moda o mediana, en este caso) de la distribución *a posteriori* es equivalente al estimador SG de (22).

Para el segundo caso (estimador A, MIDAS), se puede asumir una distribución *a priori* de β Normal con media $A\alpha$ y varianza $\sigma_\omega^2 I \rightarrow 0$ (límite no informativo) (Ullah y Raj 1980). Esto generaría una distribución *a posteriori* Normal con media equivalente al estimador A y varianza igual a $\sigma_\varepsilon^2(X'X)^{-1} - \sigma_\varepsilon^2(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R(X'X)^{-1}$.

Finalmente, para el tercer caso (estimador MCO, U-MIDAS), se puede plantear un *prior* de Jeffrey para β ($p(\beta) \propto 1$) y σ_ε^2 , así como independencia entre ambos, lo cual brindaría tras el Teorema de Bayes, una distribución *a posteriori* también bien comportada, con media igual al estimador MCO y varianza igual a $\sigma_\varepsilon^2(X'X)^{-1}$ ²⁰.

Adicionalmente, a partir de (Chib 1995), se conoce que los criterios de información típicos (Akaike y Schwarz) constituyen el límite, en muestras grandes, de la verosimilitud marginal de la data dado el modelo de regresión. Por lo tanto, se esperaría que en muestras pequeñas, una

²⁰ Se puede observar que, de los tres estimadores relevantes para la investigación (SG, A y MCO), solo el estimador SG parte de un *prior* informativo. El estimador A es una implicancia de un límite no informativo, mientras que el estimador MCO es una implicancia de un *prior* de ignorancia (no informativo). Esto guarda relación con el hecho de que la metodología FB-MIDAS respeta la incertidumbre sobre el parámetro de "suavizamiento" λ , mientras que las metodologías MIDAS y U-MIDAS proponen un tratamiento determinístico para este (lo imponen): MIDAS plantea $\lambda \rightarrow \infty$ mientras que U-MIDAS plantea $\lambda \rightarrow 0$.

metodología que use la verosimilitud marginal de la data como ponderador de modelos (BMA más adelante) o como variable a optimizar para encontrar el modelo más parsimonioso (BMS más adelante) tenga un menor error cuadrático medio de predicción (ECMP más adelante) que una metodología que utilice aquellos criterios de información. Una implicancia de esta hipótesis es que una metodología FB-MIDAS sería más eficiente que una metodología F-MIDAS, como se mencionó anteriormente.

6. Propiedades en muestras pequeñas

Los tres estimadores (SG, A y MCO) son insesgados si sus respectivos modelos están bien especificados. MCO es insesgado bajo el supuesto de exogeneidad de los regresores (Rao, Toutenburg, Shalabh y Heumann 2008).

En el caso del estimador SG en (22), se puede observar que, si $E(\omega) = 0$ (lo cual es supuesto), $E(Rb_{SG}) = 0$ (la restricción estocástica se cumple, en esperanza, con el estimador SG), con lo cual a partir de (22):

$$E(b_{SG}) = \beta + (X'X)^{-1}R'(\mu RR' + R(X'X)^{-1}R')^{-1}(0 - R\beta) = \beta$$

En el caso del estimador A en (19), también se puede ver que, si la restricción determinística se cumple $Rb_A = 0$ (lo cual siempre se da), $E(Rb_A) = 0$ (la restricción determinística se cumple, en esperanza, con el estimador A), con lo cual a partir de (19):

$$E(b_A) = \beta + (X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}(0 - R\beta) = \beta$$

Sin embargo, los tres estimadores pueden presentar sesgo de variable omitida (correlacionada con otras variables independientes) si es que el verdadero modelo correctamente especificado, que relaciona a las variables X e Y , incluye a estos. De ser el caso, el estimador A al provenir de un modelo incluido²¹ en los otros dos, sería el más sesgado. El segundo estimador más sesgado sería SG, mientras que el estimador MCO sería el menos sesgado.

Por otro lado, los estimadores A y SG tienen menores varianzas que el estimador MCO. La varianza del estimador A es:

²¹ Un modelo que incluye a otro incluye al menos las mismas variables independientes que el otro (el incluido). Un modelo no incluido en otro puede tener diferentes variables independientes que el otro (incluso con diferente cantidad de variables).

$$V(b_A) = \sigma_\varepsilon^2 (X'X)^{-1} - \sigma_\varepsilon^2 (X'X)^{-1} R' (R(X'X)^{-1} R')^{-1} R (X'X)^{-1}$$

La varianza del estimador SG es:

$$V(b_{SG}) = \sigma_\varepsilon^2 (X'X + \lambda R' (R R')^{-1} R)^{-1}$$

Finalmente, la varianza del estimador MCO es:

$$V(b_{MCO}) = \sigma_\varepsilon^2 (X'X)^{-1}$$

Se puede observar que:

$$V(b_{MCO}) - V(b_A) = \sigma_\varepsilon^2 (X'X)^{-1} R' (R(X'X)^{-1} R')^{-1} R (X'X)^{-1} \geq 0$$

Adicionalmente, si se usa el siguiente resultado:

$$(X'X + \mu R)^{-1} = (X'X)^{-1} - (X'X)^{-1} R' (\mu R R' + R(X'X)^{-1} R')^{-1} R (X'X)^{-1}$$

se puede encontrar que:

$$V(b_{MCO}) - V(b_{SG}) = \sigma_\varepsilon^2 (X'X)^{-1} R' (\mu R R' + R(X'X)^{-1} R')^{-1} R (X'X)^{-1} \geq 0$$

Finalmente, dados los resultados anteriores, y dado que $\mu R R' \geq 0$, uno puede inferir que (Rao, Toutenburg, Shalabh, & Heumann, 2008):

$$V(b_{SG}) - V(b_A) \geq 0$$

7. Error cuadrático medio de predicción (ECMP)

El error cuadrático medio de predicción (ECMP)²², según el anexo 1, se puede transformar tal que:

$$ECMP(\hat{Y}, Y) = V(\hat{Y}) + \left(\text{Sesgo}(\hat{Y}, Y) \right)^2$$

²² La definición de este concepto se puede encontrar más adelante.

La idea central de la investigación es encontrar aquel estimador que genera un menor ECMP bajo diferentes PGD. Aquel sería entonces el estimador A (metodología MIDAS), ya que tiene la menor varianza y ninguno de los tres estimadores es sesgado si se encuentran bien especificados.

Sin embargo, el modelo MIDAS es el que impone las restricciones más fuertes, con lo cual, de estar mal especificado, sería el más sesgado. Entonces, por un lado se tiene a un modelo MIDAS bastante eficiente pero posiblemente muy sesgado, y por otro lado a un modelo U-MIDAS poco sesgado pero posiblemente muy ineficiente. El objetivo de la investigación es, entonces, formular una metodología intermedia FB-MIDAS (con restricciones estocásticas), que proponga un estimador (SG) cuya ganancia en eficiencia sobre U-MIDAS más que compense el sesgo que genera, y cuyo menor sesgo que MIDAS más que compense la menor eficiencia que presenta.

Capítulo III. Metodología

1. Elección de modelos

a. *Bayesian Model Selection (BMS): calibración bayesiana empírica*

El estimador b_{SG} depende de los parámetros p (el número de rezagos autorregresivos de baja frecuencia), q (el número de rezagos de la variable independiente de alta frecuencia), d (el grado del polinomio de Almon) y λ (el parámetro de “suavizamiento”); mientras que los estimadores b_A y b_{MCO} dependen de los parámetros $\lambda = \infty, p, q, d$ y $\lambda = 0, p, q$, respectivamente. Sin embargo, una dificultad radica en que modelos asociados con diferentes combinaciones pueden ser no incluidos. Por ejemplo, es difícil elegir (para el caso FB-MIDAS) entre un modelo que usa (λ, p, q, d) con otro que emplea $(\lambda, p + 1, q, d)$; o (para el caso U-MIDAS) elegir entre un modelo que contiene $(\lambda = 0, p, q + 2)$ con otro que utiliza $(\lambda = 0, p + 1, q)$. Más aún, una limitación práctica de la estimación mixta es la determinación de λ .

Por estas razones, la investigación propone emplear una calibración bayesiana empírica para elegir el modelo que maximiza la verosimilitud marginal de la data²³ dados todos sus parámetros (la derivación de esta función se puede observar en el anexo 2) (Koop 2003).

Es decir, sea $p(Y|\lambda, p, q, d)$, la verosimilitud marginal del modelo, esta técnica propone (para la metodología FB-MIDAS, por ejemplo):

$$(\lambda_{FB}^*, p_{FB}^*, q_{FB}^*, d_{FB}^*) = \underset{(\lambda, p, q, d)}{\operatorname{argmax}} p(Y|\lambda, p, q, d)$$

Y, por lo tanto, la predicción directa sería:

$$\hat{y}_S = x_S b(\lambda_{FB}^*, p_{FB}^*, q_{FB}^*, d_{FB}^*)$$

b. *Bayesian Model Averaging (BMA): promedio ponderado de modelos bayesianos*

Sin embargo, una extensa literatura plantea que promediar modelos puede ser una mejor práctica, sobre todo cuando se tiene el propósito de predecir. Esta metodología considera la incertidumbre que uno tiene sobre la especificación de un modelo.

²³ Se utiliza la verosimilitud marginal de la data porque es el límite del criterio de información *Schwarz* (Sala-i-Martin, Doppelhoffer y Miller 2004), por lo tanto, es preferible (ambos proponen modelos parsimoniosos dado el *trade-off* entre sesgo y varianza).

Si se cuenta con \mathcal{R} diferentes modelos, y cada uno es caracterizado (de manera condicional) por una función de distribución *a priori* $p(\beta_r|\mathcal{M}_r)$, una función de verosimilitud $p(Y|\beta_r, \mathcal{M}_r)$, y finalmente, por una función de distribución *a posteriori* $p(\beta_r|Y, \mathcal{M}_r)$, entonces la inferencia multimodelo de β se basa en la distribución *a posteriori* (Koop 2003).

$$p(\beta|Y) = \sum_{r=1}^{\mathcal{R}} p(\beta_r|Y, \mathcal{M}_r)p(\mathcal{M}_r|Y)$$

donde $p(\mathcal{M}_r|Y)$ es la probabilidad *a posteriori* para el modelo r (la cual es, en sentido estricto, la probabilidad *a posteriori* (no condicional) de que el modelo r sea el verdadero PGD).

Del Teorema de Bayes, resulta que dicha probabilidad *a posteriori* satisface:

$$p(\mathcal{M}_r|Y) = Cp(Y|\mathcal{M}_r)p(\mathcal{M}_r)$$

donde $p(\mathcal{M}_r)$ es la probabilidad *a priori* del modelo r , $p(Y|\mathcal{M}_r)$ es la función de verosimilitud marginal de ese modelo, y C es una constante de integración común para todos los modelos (pues no depende de r).

Ya que el espacio de modelos es discreto²⁴, determinar C es sencillo, ya que $\sum_{r=1}^{\mathcal{R}} p(\mathcal{M}_r|Y) = 1$. Así:

$$p(\mathcal{M}_r|Y) = \frac{p(Y|\mathcal{M}_r) p(\mathcal{M}_r)}{\sum_{r=1}^{\mathcal{R}} p(Y|\mathcal{M}_r) p(\mathcal{M}_r)}$$

En términos de la investigación, se propone que la probabilidad *a priori* del modelo r sea la misma para todos los modelos²⁵, con lo cual, sin pérdida de generalidad:

$$p(\mathcal{M}_r|Y) = \frac{p(Y|\mathcal{M}_r)}{\sum_{r=1}^{\mathcal{R}} p(Y|\mathcal{M}_r)}$$

²⁴ Para la investigación, además, \mathcal{R} no es un número tan grande, por lo que no es complicado evaluar la probabilidad *a posteriori* de los modelos (de lo contrario, se tendría que usar algún algoritmo de muestreo como el *Montecarlo Markov Chain Model Composition (MC³)*).

²⁵ Todos los modelos tienen la misma probabilidad de ser elegidos, lo que representaría un *prior* de ignorancia sobre la distribución de modelos. Esto puede debatirse de acuerdo con lo establecido por Ley y Steel (2009).

Dado $L(\lambda, p, q, d)$ a partir del anexo 2, se puede hallar $p(\mathcal{M}_r|Y)$ de la siguiente manera:

$$w_r = p(\mathcal{M}_r|Y) = \frac{e^{\left(-\frac{1}{2}L_r\right)}}{\sum_{r=1}^{\mathcal{R}} e^{\left(-\frac{1}{2}L_r\right)}}$$

tal que, si el modelo r genera una predicción $\widehat{y}_{rs} = x_s b_r$, se puede mejorar la inferencia con una predicción combinada²⁶ como

$$\widehat{y}_s = \sum_{r=1}^{\mathcal{R}} w_r \widehat{y}_{rs}$$

Incluso, BMA puede ser usado para análisis multivariado. Si se permite que X_s dependa también de r , no solo se pueden encontrar predicciones combinadas de diferentes especificaciones dado un indicador X_s , sino que también una predicción combinada de diferentes indicadores.

2. Especificación de modelos

a. Especificación condicional

Caso BMS:

Elegir de manera *ex ante* y *ad hoc* las variables $p = p_0, q = q_0, d = d_0$ que serán comunes para las tres metodologías.

Luego, para FB-MIDAS elegir $(\lambda^*, p = p_0, q = q_0, d = d_0)$ tal que:

$$\lambda^* = \underset{\lambda}{\operatorname{argmin}} [L(\lambda, p, q, d) \mid p = p_0, q = q_0, d = d_0]^{27}$$

Para MIDAS, elegir aquel modelo con $(\lambda = \infty, p = p_0, q = q_0, d = d_0)$. Finalmente, para U-MIDAS, elegir aquel modelo con $(\lambda = 0, p = p_0, q = q_0)$.

²⁶ Es fácil demostrar que la varianza estimada de $\widehat{y}_s = \sum_{r=1}^{\mathcal{R}} w_r x_s b_r$ (condicional en la información histórica de X_s), es igual a $v_r = \widehat{V}_0 r_{rs} r'_{rs}$, donde $r_{rs} = X_s (X'_s X_s + \lambda_r R'_r (R_r R'_r)^{-1} X'_s)$, y $\widehat{V}_0 = Y' M Y / \operatorname{tr}(M)$ para $M = I_n - X(X'X)^{-1}X'$ es el estimador MCO insesgado de la varianza del error. Por otro lado, (Burnham y Anderson 2002) han propuesto:

$$v_y = \left(\sum_{r=1}^{\mathcal{R}} w_r \sqrt{v_r + (\widehat{y}_{rs} - \widehat{y}_s)^2} \right)^2$$

como un estimador de la varianza de \widehat{y}_s , el cual toma en cuenta tanto la variabilidad muestral como la incertidumbre del modelo.

²⁷ Donde $L(\lambda, p, q, d)$ se puede observar en el anexo 2.

En cualquiera de las tres metodologías, la predicción es directa: Para FB-MIDAS sería $y_{FB,S}^{\widehat{BMS},C} = x_S b_{(\lambda^*, p=p_0, q=q_0, d=d_0)}$; en MIDAS sería $y_{M,S}^{\widehat{BMS},C} = x_S b_{(\lambda=\infty, p=p_0, q=q_0, d=d_0)}$; Y finalmente, para el caso U-MIDAS sería $y_{U,S}^{\widehat{BMS},C} = x_S b_{(\lambda=0, p=p_0, q=q_0)}$.

Caso BMA:

También elegir de manera *ex ante* y *ad hoc* las variables $p = p_0, q = q_0, d = d_0$ que serán comunes para las tres metodologías, y promediar modelos a lo largo de la verosimilitud marginal de la data (con ponderaciones de verosimilitud marginal).

Luego, para FB-MIDAS proponer diferentes modelos $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3, \dots, \mathcal{M}_R$, cada uno con diferentes valores de λ , tal que $y_{FB,S}^{\widehat{BMA},C} = \sum_{r=1}^R w_r x_S b_r$. Uno se puede dar cuenta que, para MIDAS y U-MIDAS, como todos los valores son impuestos (y no existiría una lista de valores de parámetros para promediar), las predicciones serán equivalentes a las que se obtienen bajo la metodología BMS: $y_{M,S}^{\widehat{BMA},C} = y_{M,S}^{\widehat{BMS},C}$ y $y_{U,S}^{\widehat{BMA},C} = y_{U,S}^{\widehat{BMS},C}$, respectivamente.

b. Especificación no condicional

Caso BMS:

Para cada metodología, obtener el modelo óptimo maximizando la verosimilitud marginal de la data.

De esta manera, para FB-MIDAS elegir $(\lambda_{FB}^*, p_{FB}^*, q_{FB}^*, d_{FB}^*)$, donde:

$$(\lambda_{FB}^*, p_{FB}^*, q_{FB}^*, d_{FB}^*) = \underset{(\lambda, p, q, d)}{\operatorname{argmin}} L(\lambda, p, q, d)$$

Para MIDAS elegir $(\lambda = \infty, p_M^*, q_M^*, d_M^*)$, donde:

$$(p_M^*, q_M^*, d_M^*) = \underset{(p, q, d)}{\operatorname{argmin}} [L(\lambda, p, q, d) \mid \lambda = \infty]$$

Finalmente, para U-MIDAS elegir $(\lambda = 0, p_U^*, q_U^*)$, donde:

$$(p_U^*, q_U^*) = \underset{(p, q)}{\operatorname{argmin}} [L(\lambda, p, q, d) \mid \lambda = 0]$$

De igual manera, la predicción es directa en cualquiera de las tres metodologías: Para FB-MIDAS

sería $y_{FB,S}^{\widehat{BMS},N} = x_s b_{(\lambda_{FB}^*, p_{FB}^*, q_{FB}^*, d_{FB}^*)}$; en MIDAS sería $y_{M,S}^{\widehat{BMS},N} = x_s b_{(\lambda=\infty, p_M^*, q_M^*, d_M^*)}$; y finalmente, para U-MIDAS sería $y_{U,S}^{\widehat{BMS},N} = x_s b_{(\lambda=0, p_U^*, q_U^*)}$.

Caso BMA:

Para cada metodología, promediar modelos a lo largo de la verosimilitud marginal de la data.

De esta manera, para el caso FB-MIDAS proponer diferentes modelos $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3, \dots, \mathcal{M}_{\mathcal{R}}$, cada uno con diferentes combinaciones de valores de λ, p, q y d , tal que $y_{FB,S}^{\widehat{BMA},N} = \sum_{r=1}^{\mathcal{R}} w_r x_s b_r$. En cuanto a MIDAS, proponer diferentes modelos con combinaciones diferentes de $\lambda = \infty, p, q$ y d , tal que $y_{M,S}^{\widehat{BMA},N} = \sum_{r=1}^{\mathcal{R}} w_r x_s b_r$. Finalmente, para U-MIDAS, proponer diferentes modelos con combinaciones diferentes de $\lambda = 0, p$ y q , tal que $y_{U,S}^{\widehat{BMA},N} = \sum_{r=1}^{\mathcal{R}} w_r x_s b_r$ ²⁸.

3. Comparación de modelos

a. Ratio de medias de ECMP

Para comparar FB-MIDAS con MIDAS y FB-MIDAS con U-MIDAS, se calcula el ECMP para cada metodología de selección. El ECMP se define, a partir del anexo 1, de la siguiente manera:

$$ECMP(\hat{Y}, Y) = E(\hat{Y} - Y)^2$$

La investigación emplea una técnica recursiva y un número dado de simulaciones de Montecarlo, que permitirán capturar un promedio (que por Ley de Grandes Números (LGN) tiende a la esperanza) de ECMP para cada metodología.

En particular, sea I el número de iteraciones recursivas y S el número de simulaciones de Montecarlo:

$$\frac{1}{I} \sum_{i=1}^I (Y_i - \hat{Y}_i)^2 = ECMP(\hat{Y}_i, Y)$$

Por lo tanto:

²⁸ Se pueden observar más combinaciones (y por lo tanto más modelos) en el caso de FB-MIDAS. En el caso MIDAS hay menos combinaciones pues se fija $\lambda = \infty$, mientras que en el caso U-MIDAS hay aún menos combinaciones pues se fija $\lambda = 0$ y el valor de d es irrelevante para todo fin práctico.

$$\frac{1}{SI} \sum_{s=1}^S \sum_{i=1}^I (Y_{is} - \hat{Y}_{is})^2 \quad \xrightarrow{p} \quad E[ECMP(\hat{Y}_i, Y)]$$

que no es otra cosa más que la esperanza del ECMP para una metodología particular i .

Una vez que se obtiene este resultado, se halla el ratio de medias de ECMP (REECMP) entre FB-MIDAS y MIDAS y entre FB-MIDAS y U-MIDAS²⁹:

$$\frac{\frac{1}{SI} \sum_{s=1}^S \sum_{i=1}^I (Y_{is} - \hat{Y}_{is}^{FB})^2}{\frac{1}{SI} \sum_{s=1}^S \sum_{i=1}^I (Y_{is} - \hat{Y}_{is}^k)^2} \quad \text{para } k$$

$$= MIDAS, U - MIDAS$$

b. Ratio de percentiles de ECMP

Para hacer más robusto el análisis y tomar en cuenta la forma de la distribución del ECMP de cada metodología, no solo se propone analizar los ratios de medias sino también los ratios de percentiles 25, 50 (mediana), y 75 (RPECMP):

$$\frac{P_m[ECMP(\hat{Y}_{FB})]}{P_m[ECMP(\hat{Y}_k)]} \quad \text{para } k = MIDAS, U - MIDAS \text{ y } m = 25, 50, 75$$

²⁹ Foroni, Marcellino y Schumacher (2015) y Winkelried (2012) computan la media del ratio de ECMP, a diferencia del ratio de medias de ECMP (como en esta especificación). Proponer la media del ratio de ECMP no ayudaría en el objetivo de esta investigación, ya que se busca obtener una comparación sin distorsiones.

Capítulo IV. Ejercicio de evaluación

1. Procesos generadores de datos (PGD)

Ya que en series de tiempo es difícil obtener un *ranking* solo tomando en cuenta consideraciones teóricas³⁰, se realizan simulaciones de Montecarlo con diferentes procesos generadores de datos (PGD) que permiten resaltar las ventajas de FB-MIDAS sobre MIDAS y U-MIDAS de manera robusta. En particular, se utilizan 3 PGD³¹: VAR (1), MIDAS PGD y U-MIDAS PGD.

a. VAR (1)

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \rho & \delta_l \\ \delta_h & \rho \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} e_{y,t} \\ e_{x,t} \end{pmatrix}$$

donde y_t es la variable de baja frecuencia y x_t es la variable de alta frecuencia³². Además, $e_{y,t}$ y $e_{x,t}$ siguen una distribución Normal con media cero y varianza tal que la varianza no condicional de y_t es igual a 1. Para el parámetro de persistencia se escoge $\rho = \{0,1, 0,5, 0,9\}$, y adicionalmente, se fijan δ_l y δ_h tal que $\delta_l = \{0,1, 0,5, 1,0\}$ y $\delta_h = 0$. Por lo cual, el PGD es recursivo.

b. MIDAS PGD

$$\gamma(L^m)y_t = \beta_1 B(L; \theta)x_t + \varepsilon_t$$

En este segundo caso, se plantea un PGD a partir del modelo de las ecuaciones (3) y (4). En particular, se propone $dd = 2$, $p = 1$, $\theta_1 = 0,7$, $\theta_2 = \{-0,03, -0,05, -0,3, -0,2, -0,1\}$, y finalmente, x_t sigue un proceso AR (1) (es decir, $q = 1$) con persistencia igual a 0,9 (igual a γ).

c. U-MIDAS PGD

$$\gamma(L^m)y_t = \beta(L)x_t + \varepsilon_t$$

Finalmente, se plantea un PGD a partir del modelo en la ecuación (5), y se mantienen los mismos parámetros (relevantes para este caso) de la especificación anterior: $p = q = 1$ y $\gamma = \phi = 0,9$.

³⁰ Sus méritos relativos dependen muy fuertemente del PGD.

³¹ Extraídos de Foroni, Marcellino y Schumacher (2015). En cada PGD, hay diferentes combinaciones de parámetros.

³² δ_l aplica para la variable de baja frecuencia (*low*), mientras que δ_h aplica para la variable de alta frecuencia (*high*).

2. Data empírica

Para aterrizar el análisis de robustez, se plantea utilizar la metodología de la investigación en predicciones sobre variables macroeconómicas relevantes.

Para la disparidad trimestral-mensual, se propone un modelo de PBI real vs. algunos indicadores líderes conocidos: importaciones de bienes de capital sin materiales de construcción, expectativas de la economía del BCRP a 3 meses, consumo interno de cemento, producción de electricidad, desembarque de anchoveta y empleo de Lima Metropolitana³³.

Para la disparidad trimestral-diaria, se utiliza o bien un modelo de PBI real o uno de consumo real, pero esta vez vs. algunas variables conocidas disponibles en alta frecuencia: tipo de cambio, spread EMBIG, Bolsa de Valores, precio del cobre y producción de electricidad³⁴.

Finalmente, para la disparidad mensual-diaria, se emplea o bien un modelo de inflación de alimentos (mensual) vs. el precio del pollo, la papa y el arroz (disponibles de manera diaria)³⁵, o bien un modelo de PBI real mensual vs. los indicadores líderes diarios ya presentados. Se puede observar en el anexo 3 las predicciones de las tres variables macroeconómicas presentadas (una para cada disparidad), combinando diferentes elecciones de modelos, especificaciones de modelos, h y k .

3. Estrategia empírica

La base de datos, para cada PGD (y cada combinación de parámetros de cada PGD), cuenta con 72 valores de la variable de baja frecuencia trimestral y 216 de la variable de baja frecuencia mensual. De esta manera, también se cuenta con $72 \times 3 + 3 = 219$ valores de la variable de alta frecuencia (mensual, para disparidad trimestral-mensual), $72 \times 90 + 30 = 6.510$ valores de la variable de alta frecuencia (diaria, para disparidad trimestral-diaria), y $216 \times 30 + 10 = 6.490$ valores de la variable de alta frecuencia (diaria, para disparidad mensual-diaria)³⁶. La base de datos de data empírica consta de 60 períodos de variable dependiente trimestral y 120 períodos de variable dependiente mensual.

³³ Estas variables han sido encontradas tanto en la página web del Banco Central de Reserva del Perú (BCRP) como en la página web del Instituto Nacional de Estadística e Informática (INEI), en donde también se encuentra la data de actividad económica mensual (PBI real mensual en un párrafo posterior).

³⁴ También disponibles en BCRP, INEI o Bloomberg.

³⁵ Datos disponibles en el MINAGRI.

³⁶ La existencia de 3 (o 30 o 10 según sea el caso) datos adicionales de la variable de alta frecuencia se debe a que son necesarios para hacer un análisis de *backcasting*.

Se consideran tres casos para *nowcasting* $h = \{-2, -1, 0\}$ y tres casos para *backcasting* $h = \{1, 2, 3\}$ ³⁷, para situaciones tanto cuando la observación trimestral inmediata anterior está disponible ($\kappa = 0$) como cuando hay un rezago en publicación de un trimestre ($\kappa = 1$).

Como se mencionó, de manera ligera previamente, para la data empírica se utiliza un esquema de estimación recursivo: la muestra consiste de una ventana móvil de 42 trimestres, con lo cual se cuenta con 30 ventanas ($I = 30$).

Adicionalmente, para cada combinación de parámetros de cada PGD (3 por cada uno, cuyos resultados al final son promediados) se realizan 10.000 simulaciones ($S = 1.000$).

Se obtienen, entonces, 17.280.000 numeradores y 17.280.000 denominadores de ECMP (3 PGD x 3 combinaciones de parámetros por PGD x 6 valores de h x 2 valores de k x 2 metodologías de elección de modelos³⁸ x 2 metodologías de especificación de modelos³⁹ x 4 metodologías de estimación x 10.000 simulaciones). Finalmente, al realizar las metodologías de comparación de modelos (4 estadísticas: media, mediana, y P50) se obtienen los 9.600 ratios que constituirán la base del análisis del trabajo.

³⁷ Dependiendo de la disparidad, el *nowcasting* y *backcasting* varía.

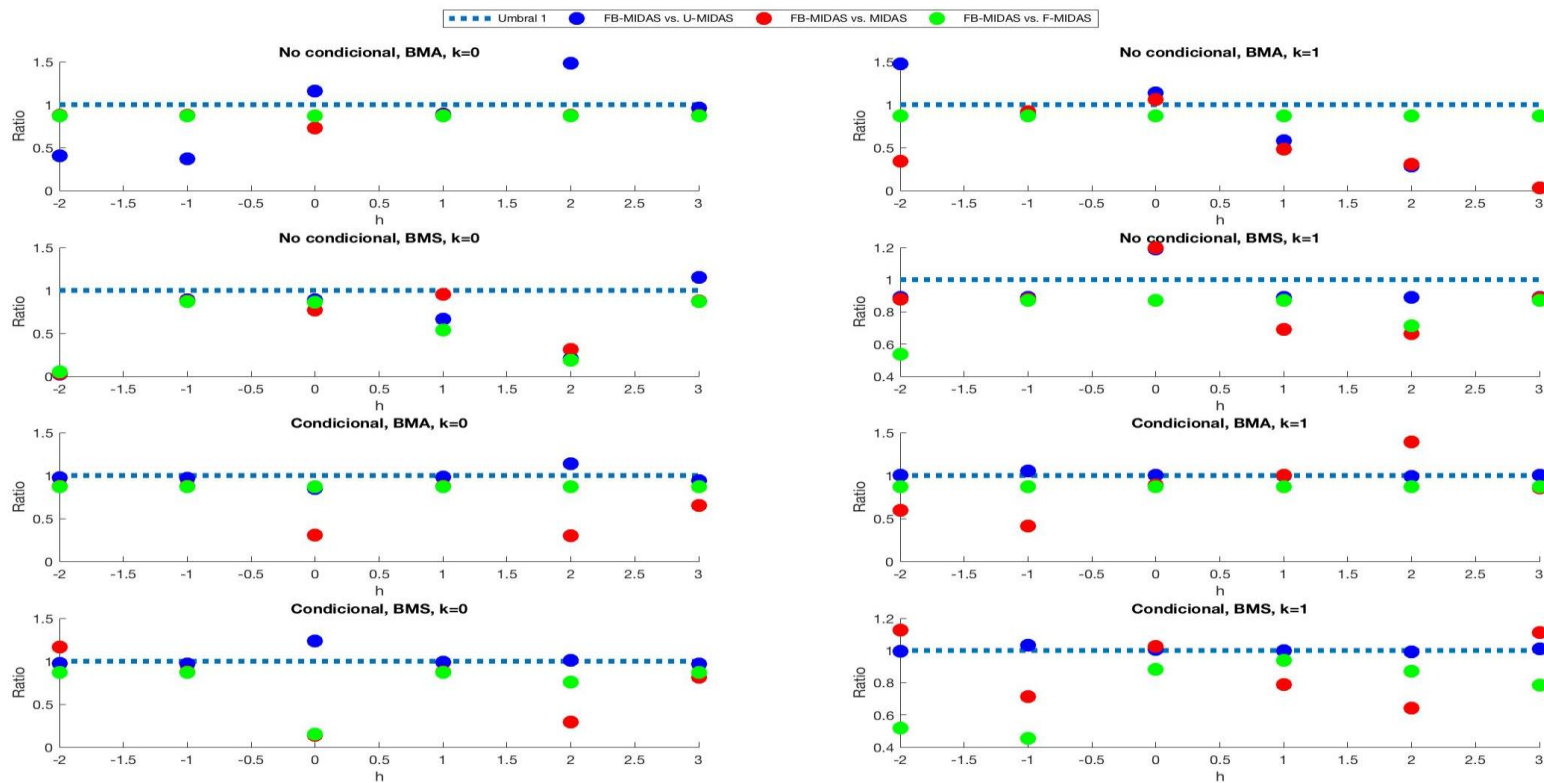
³⁸ Respecto a BMA, se plantean los siguientes valores: $p = \{1, 2\}$, $q = \{3, 6, 9, 12\}$, $d = \{1, 2, 3, 4\}$, $\lambda = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ (se observó que λ fluctúa a través de estos valores a lo largo de 30 iteraciones en 5 simulaciones prueba, cuando se realizó BMS).

³⁹ Respecto a la especificación condicional, se plantean los siguientes valores: $p = 2$, $q = 8$, $d = 2$.

Capítulo V. Análisis de resultados

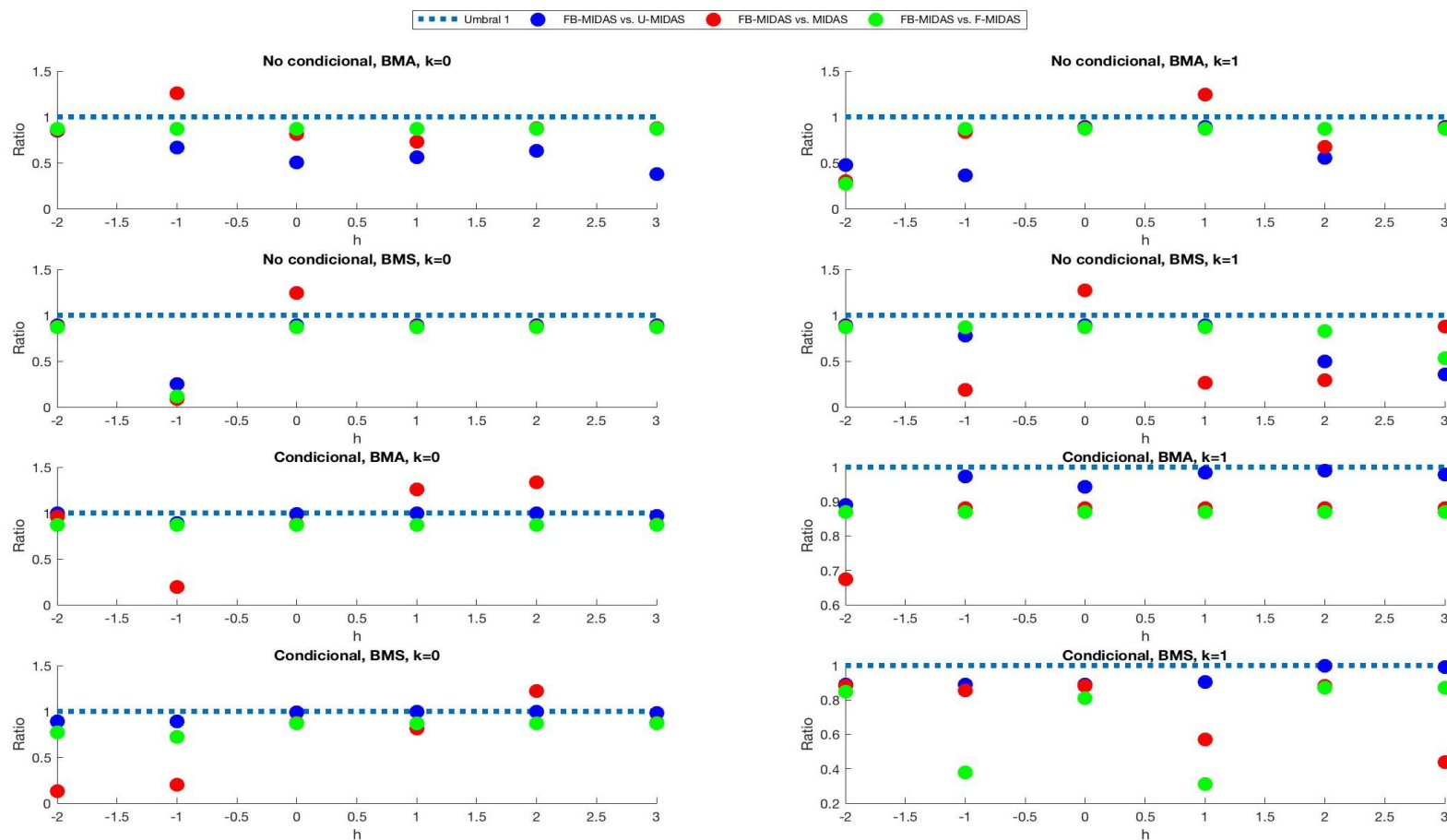
1. Disparidad trimestral-mensual

Gráfico 1. VAR (1)



El gráfico muestra el REECMP y RPECMP para diferentes combinaciones de h y k , las dos metodologías de elección de modelos, las dos metodologías de especificación de modelos, y diferentes especificaciones del PGD VAR (1). Elaboración propia a través de Matlab ®.

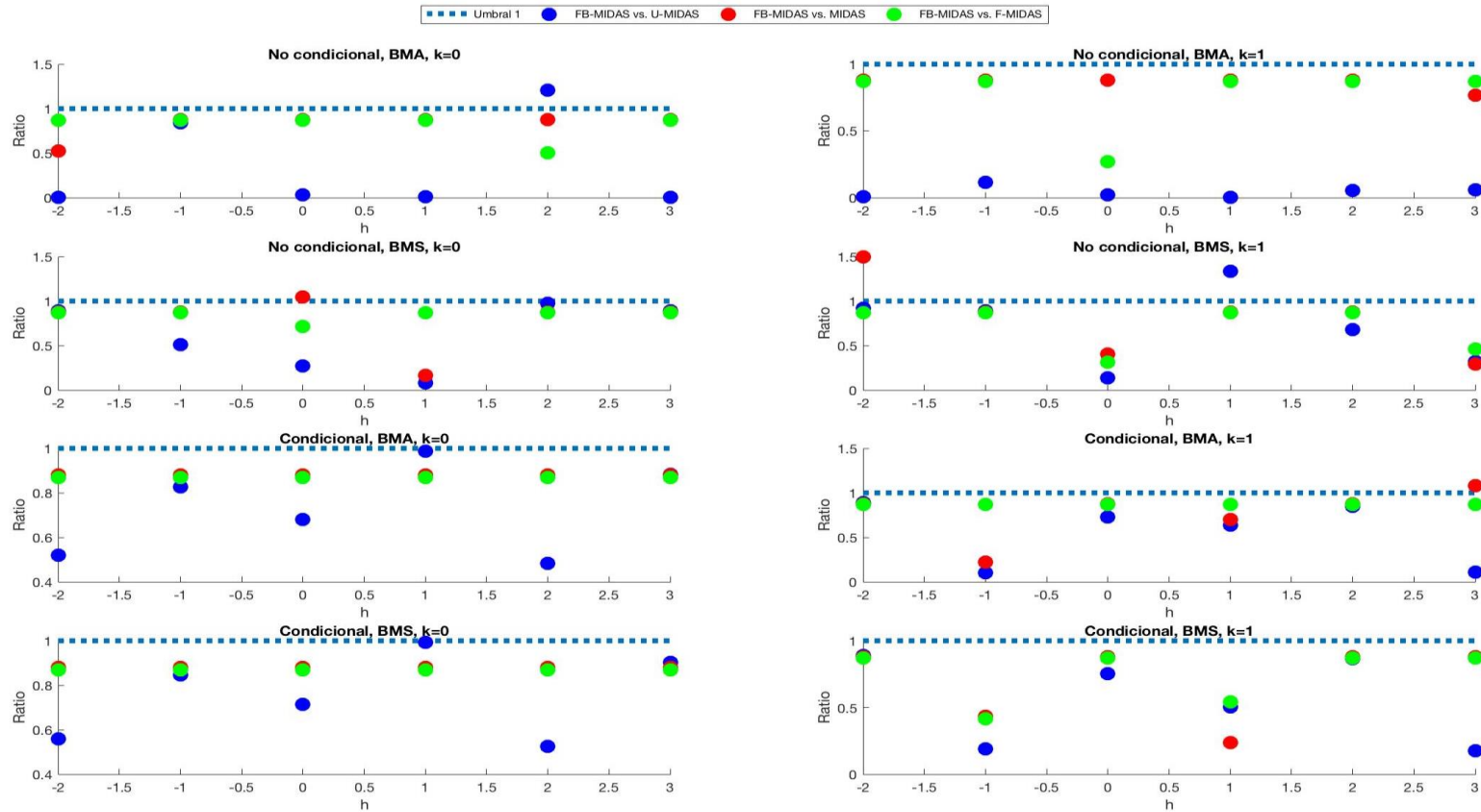
Gráfico 2. PBI real trimestral vs. indicadores líderes mensuales



El gráfico muestra el REECMP y RPECMP para diferentes combinaciones de h y κ , las dos metodologías de elección de modelos, las dos metodologías de especificación de modelos de la data empírica del PBI real trimestral vs. sus indicadores líderes mensuales. Elaboración propia a través de Matlab ®.

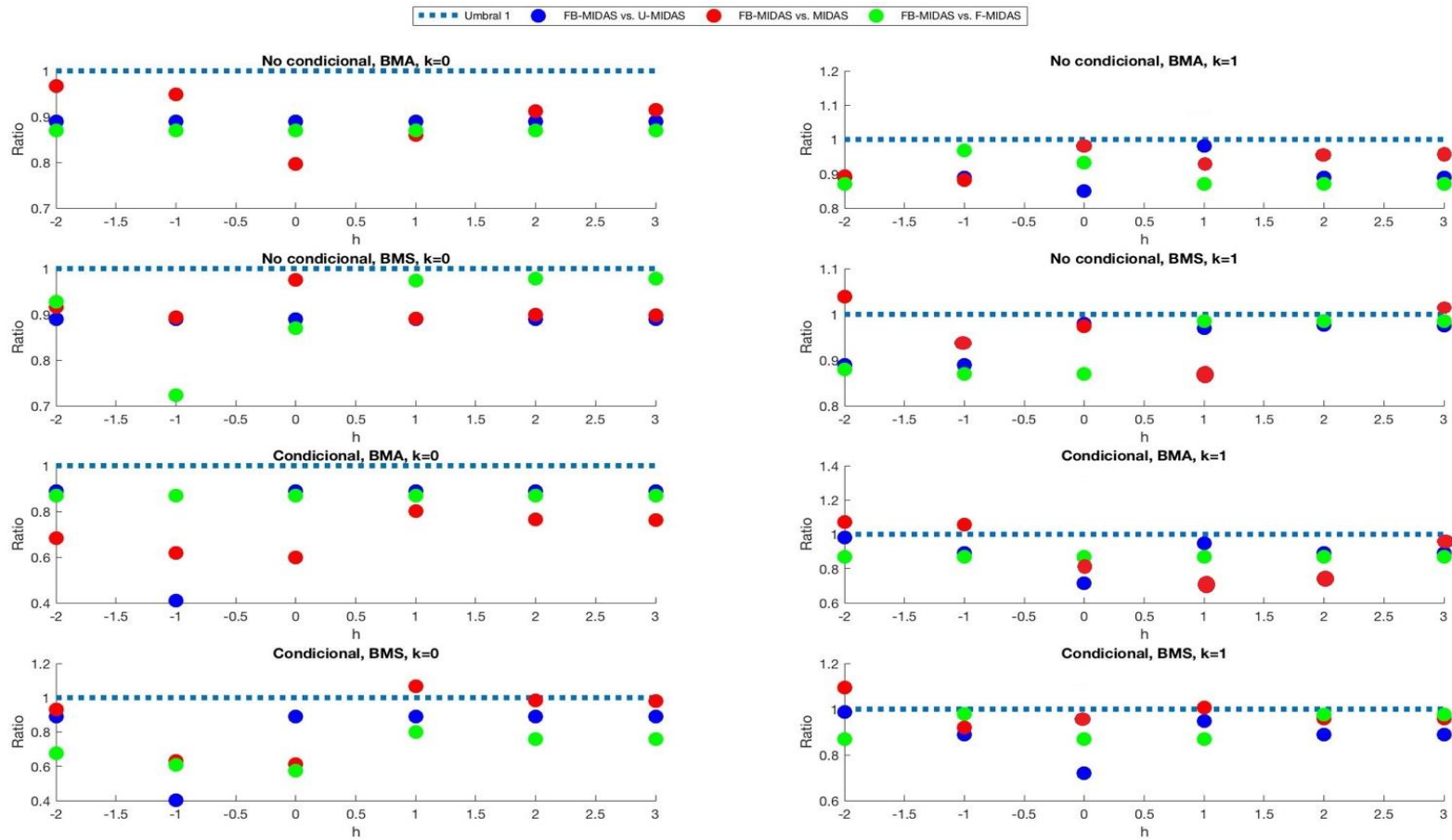
2. Disparidad trimestral-diaria

Gráfico 3. MIDAS PGD



El gráfico muestra el REECMP y RPECMP para diferentes combinaciones de h y κ , las dos metodologías de elección de modelos, las dos metodologías de especificación de modelos, y diferentes especificaciones del PGD MIDAS. Elaboración propia a través de Matlab®.

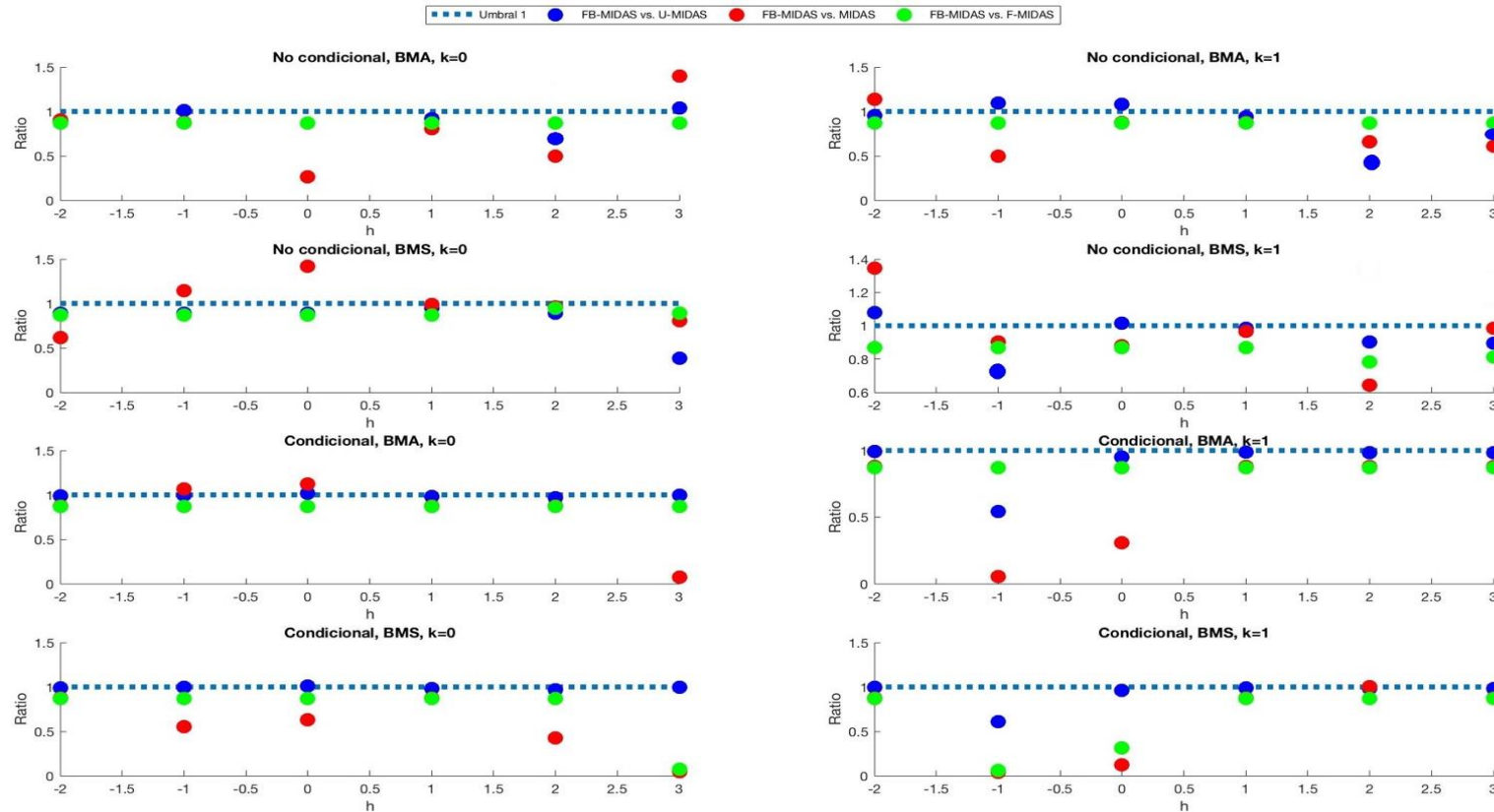
Gráfico 4. Consumo real trimestral vs. indicadores líderes diarios



El gráfico muestra el REECMP y RPECMP para diferentes combinaciones de h y κ , las dos metodologías de elección de modelos, las dos metodologías de especificación de modelos, y diferentes especificaciones de la data empírica del consumo real trimestral vs. sus indicadores líderes diarios. Elaboración propia a través de Matlab ®.

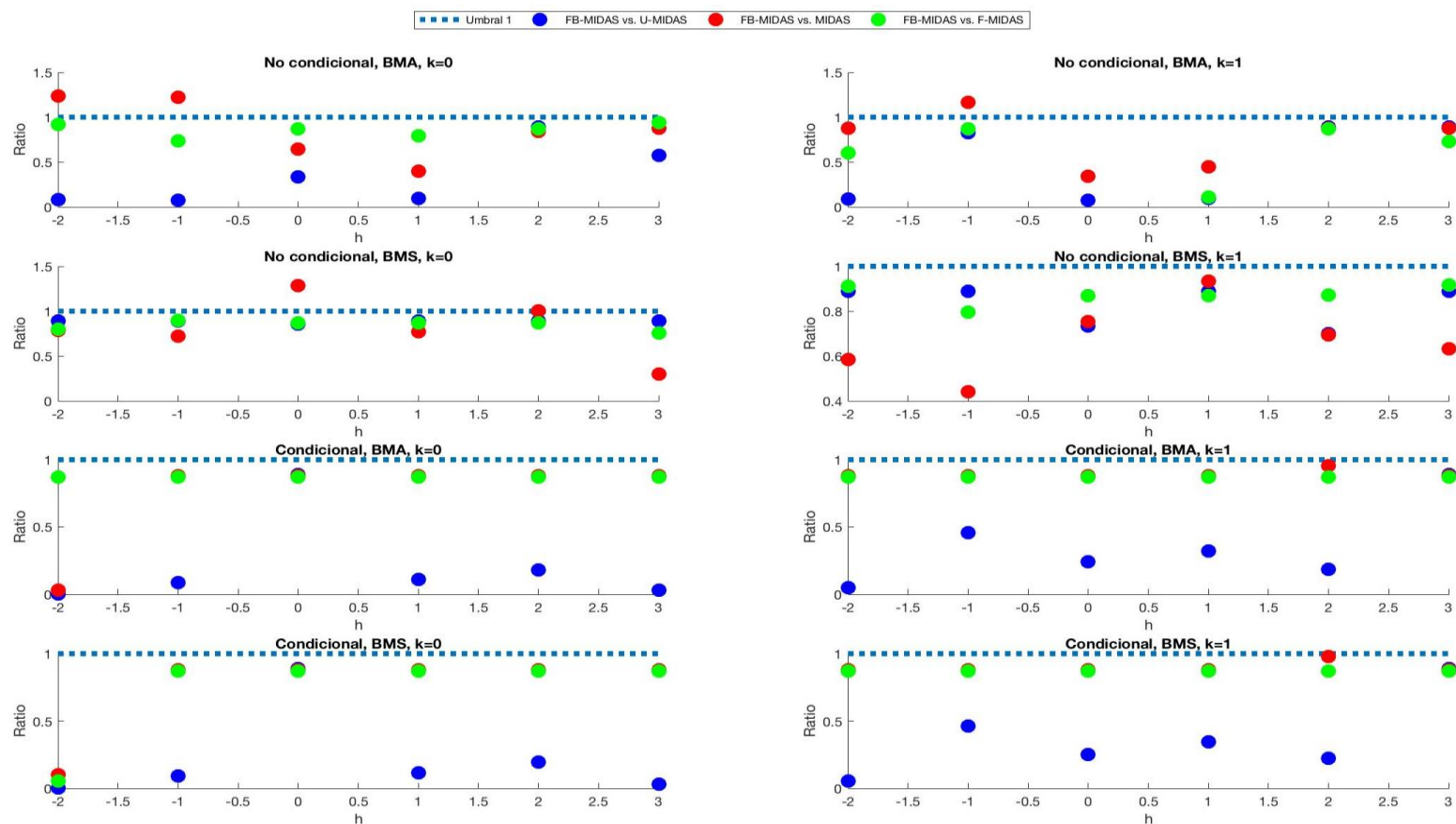
3. Disparidad mensual-diaria

Gráfico 5. U-MIDAS PGD



El gráfico muestra el REECMP y RPECMP para diferentes combinaciones de h y κ , las dos metodologías de elección de modelos, las dos metodologías de especificación de modelos, y diferentes especificaciones del PGD U-MIDAS. Elaboración propia a través de Matlab®.

Gráfico 6. PBI real mensual vs. indicadores líderes diarios



El gráfico muestra el REECMP y RPECMP para diferentes combinaciones de h y κ , las dos metodologías de elección de modelos, las dos metodologías de especificación de modelos, y diferentes especificaciones de la data empírica del PBI real mensual vs. sus indicadores líderes diarios. Elaboración propia a través de Matlab ®.

Se puede observar que, en general, los resultados varían dependiendo de: a) h o k , b) la combinación de parámetros usada dentro de cada PGD, c) la metodología de selección de modelos y d) la metodología de especificación de modelos. Sin embargo, se puede concluir que, en general, FB-MIDAS genera una mejor predicción, mediante una menor esperanza de ECMP, que MIDAS, U-MIDAS y F-MIDAS, para la disparidad de frecuencias trimestral-mensual, trimestral-diaria y mensual-diaria.

Esta conclusión es robusta al PGD que se emplea (ya sea empírico o paramétrico). Adicionalmente, aun cuando el PGD es MIDAS o U-MIDAS, FB-MIDAS resulta ser bastante competitivo. Se observa claramente que cuando el PGD es MIDAS, resulta más difícil competir vs. MIDAS (medias y medianas mayores que 1), mientras que cuando el PGD es U-MIDAS también resulta más difícil competir con U-MIDAS.

No se observan muchas diferencias entre las metodologías de elección de modelos, pero sí se observan algunas diferencias entre las metodologías de especificación de modelos presentadas. En particular, cuando la especificación es condicional a ciertos parámetros, la precisión de FB-MIDAS es menor que cuando la especificación es no condicional. Cabe recordar que la especificación no condicional busca encontrar los óptimos de cada metodología.

Incluso se puede realizar un *double-check* de esta última conclusión revisando los valores que se obtienen de lambda óptimo cuando se emplea BMS, los cuales se pueden observar en el anexo 4. Cuando la disparidad es trimestral-mensual los valores de lambda son cercanos a 0, que hace sentido puesto que un valor de lambda de 0 significa un estimador MCO de la metodología U-MIDAS, entendiéndose que este sería el competidor más potente en este caso. Cuando la disparidad es trimestral-diaria los valores de lambda son bastante altos (elevados a la sexta potencia decimal), que también hace sentido puesto que un valor de lambda de 0 significa un estimador MCNL de la metodología MIDAS, entendiéndose también que este sería el competidor más fuerte en este caso. Finalmente, cuando la disparidad es mensual-diaria, los valores de lambda son altos pero no tanto como en el caso trimestral-diario, entendiéndose que las conclusiones aquí dependerían del PGD subyacente.

Por otro lado, se puede ver que, en el peor de los casos, el ratio de medias y percentiles de ECMP llega a ser 1,4 para cualquier combinación de parámetros de cualquier PGD, con lo cual uno también podría concluir que FB-MIDAS resulta ser bastante competitivo respecto de otros modelos MIDAS determinísticos (MIDAS y U-MIDAS) y frecuentistas (F-MIDAS).

Para ahondar en el análisis, se encuentra que U-MIDAS resulta ser el competidor más fuerte cuando la disparidad entre frecuencias es de trimestral a mensual. Esto está en línea con la literatura (Froni, Marcellino y Schumacher 2015). Adicionalmente, se observa que, mientras la variable mensual esté más alejada de la variable trimestral (h es grande en valor absoluto), las predicciones FB-MIDAS empeoran, sin embargo, permanecen estables dentro del límite máximo encontrado de 1,4 (y con la mayoría de los ratios con un valor menor que el umbral de 1).

Los resultados son reveladores: cuando se mezcla data trimestral con data mensual, el *ranking* de precisión es FB-MIDAS, U-MIDAS y MIDAS. Cuando se combina data trimestral con diaria, el *ranking* es FB-MIDAS, MIDAS y U-MIDAS. Finalmente, cuando se junta data mensual con diaria, el *ranking* depende del PGD que se emplee (en particular, el gráfico anterior muestra un PGD U-MIDAS, y se observa justamente que U-MIDAS resulta ser un competidor más fuerte).

Respecto a la comparación con F-MIDAS, se observa que cuando se usa una elección de modelos BMA, los resultados son equivalentes, ya que el universo de modelos es el mismo en ambos casos. En cambio, cuando se utiliza una elección BMS, los resultados son diferentes y favorecen a FB-MIDAS, indicando su eficiencia en muestras pequeñas.

Para terminar, a partir de los resultados mostrados en el anexo 3, se comprueba la significancia estadística de la superioridad de FB-MIDAS sobre las demás metodologías.

Se concluye entonces que FB-MIDAS puede ser una buena alternativa para MIDAS, U-MIDAS y F-MIDAS, debido a que disminuye el sesgo relativo a MIDAS, la varianza relativa a U-MIDAS y es flexible para trabajar con muestras pequeñas, generando eficiencia sobre F-MIDAS.

Conclusiones y recomendaciones

1. Conclusiones

Los resultados de la investigación son reveladores: si se imponen restricciones de manera estocástica, determinadas empíricamente, se pueden encontrar casos en donde las predicciones mejoran respecto a los casos en donde o se imponen restricciones muy fuertes, o simplemente no se imponen restricciones. Esta mejora radica en el *trade-off* sesgo-varianza. Estos resultados son robustos al PGD: la investigación ha utilizado diferentes combinaciones de cada uno de los 3 PGD paramétricos y 3 PGD empíricos, y se han obtenido las mismas conclusiones, incluso con significancia estadística.

2. Recomendaciones

Clements y Galvão (2008) demuestran que se pueden plantear modelos con variables independientes de más de una frecuencia (por ejemplo, mensual y diaria). Esto podría ayudar a generar un indicador de alta frecuencia de alguna variable macroeconómica trimestral: utilizar la información de variables mensuales y diarias, por ejemplo, para tener un indicador líder potente de PBI real trimestral de manera diaria.

Bibliografía

- Akaike, H. (1979a). “Likelihood and the Bayes Procedure”. *International Meeting on Bayesian Statistics*, 1-47.
- Akaike, H. (1979b). “Smoothness Priors and the Distributed Lag Estimator”. *Stanford University*, 1-14.
- Andreou, E., Ghysels, E., & Kourtellis, A. (2010). “Regression models with mixed sampling frequencies”. *Journal of Econometrics*, 158(2), 246-261.
- Armesto, M. T., Engemann, K. M., & Owyang, M. T. (2010). “Forecasting with mixed frequencies”. *Federal Reserve Bank of St. Louis*, 92(6), 521-536.
- Bai, J., Ghysels, E., & Wright, J. (2013). “States space models and Midas regressions”. *Econometric Reviews*. Vol. 32, pp. 779-813.
- Burnham, K. P. & Anderson, D. R. (2002). *Model Selection and Multimodel Inference. A Practical Information-Theoretic Approach*. Springer-Verlag (2).
- Chib, S. (1995). “Output, Marginal Likelihood from the Gibbs”. *Journal of the American Statistical Association*, 90(432).
- Clements, M. P. & Galvão, A. B. (2008). “Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States”. *Journal of Business & Economic Studies*, 26(4), 546-554.
- Clements, M. P. & Galvão, A. B. (2009). “Forecasting US output growth using leading indicators: An appraisal using MIDAS models”. *Journal of Applied Econometrics*, 24(7), 1187-1206.
- Forni, C. & Marcellino, M. (2013). “A survey of econometric methods for mixed-frequency data”. *Norges Bank Research*, 1-45.
- Forni, C., Marcellino, M. & Schumacher, C. (2015). “Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials”. *Journal of the Royal Statistical Society*, 178, 57-82.
- Ghysels, E., Santa-Clara, P. & Valkanov, R. (2004). *The MIDAS touch: Mixed data sampling regression models*. CIRANO.
- Ghysels, E., Sinko, A. & Valkanov, R. (2007). “MIDAS Regressions: Further Results and New Directions”. *Econometric Views*, 26(1), 53-90.
- Giles, D. E. (1975). *Bayesian Applications in Econometrics*. Tesis doctoral. University of Canterbury, 1-278.
- Koop, G. (2003). *Bayesian Econometrics*. Glasgow: University of Glasgow.
- Kuzin, V., Marcellino, M. & Schumacher, C. (2011). “MIDAS vs. mixed-frequency VAR:

- Nowcasting GDP in the Euro are”. *International Journal of Forecasting*, 27(2), 529-542.
- Ley, E. & Steel, M. F. (2009). “On the effect of prior assumptions in Bayesian Model Averaging with Applications to Growth Regressions”. *Journal of Applied Econometrics*, 24, 651-674.
 - Maddala, G. S. (1974). “Ridge Estimators for Distributed Lag Models”. *NBER* (69), 1-20.
 - Proietti, T. (2006). “Temporal disaggregation by state space methods: Dynamic regression methods revisited”. *Econometrics Journal*, 9(3), 357-372.
 - Rao, C., Toutenburg, H., Shalabh, & Heumann, C. (2008). *Linear Models and Generalizations*. Berlin, Germany: Springer.
 - Sala-i-Martin, X., Doppelhoffer, G. & Miller, R. (2004). “Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach”. *The American Economic Review*, 94(4), 813-835.
 - Shiller, R. J. (1973). “A distributed lag estimator derived from smoothness priors”. *Econometrica*, 41(4), 775-788.
 - Theil, H. & Goldberger, A. S. (1961). “On Pure and Mixed Statistical Estimation in Economics”. *International Economic Review*, 2(1), 65-78.
 - Ullah, A. & Raj, B. (1979). “A distributed lag estimator derived from Shiller’s smoothness priors (An extension)”. *Economics Letters*, 2, 219-223.
 - Ullah, A. & Raj, B. (1980). “A Polynomial Distributed Lag Model with Stochastic Coefficients and Priors”. *Journal of Empirical Economics*, 5, 219-232.
 - Winkelried, D. (2012). “Predicting quarterly aggregates with monthly indicators”. *Banco Central de Reserva del Perú*, 23, 1-16.

Anexos

Anexo 1. Error cuadrático medio de predicción (ECMP)

$$ECMP(\hat{Y}, Y) = E(\hat{Y} - Y)^2$$

$$ECMP(\hat{Y}, Y) = E(\hat{Y} - E(\hat{Y}) + E(\hat{Y}) - Y)^2$$

$$ECMP(\hat{Y}, Y) = E\left(\left(\hat{Y} - E(\hat{Y})\right)^2 + 2\left(\hat{Y} - E(\hat{Y})\right)\left(E(\hat{Y}) - Y\right) + \left(E(\hat{Y}) - \hat{Y}\right)^2\right)$$

$$ECMP(\hat{Y}, Y) = E\left(\left(\hat{Y} - E(\hat{Y})\right)^2\right) + E\left(2\left(\hat{Y} - E(\hat{Y})\right)\left(E(\hat{Y}) - Y\right)\right) + E\left(\left(E(\hat{Y}) - \hat{Y}\right)^2\right)$$

$$ECMP(\hat{Y}, Y) = E\left(\left(\hat{Y} - E(\hat{Y})\right)^2\right) + 2\left(E(\hat{Y}) - \hat{Y}\right)E\left(\hat{Y} - E(\hat{Y})\right) + \left(E(\hat{Y}) - \hat{Y}\right)^2$$

$$ECMP(\hat{Y}, Y) = E\left(\left(\hat{Y} - E(\hat{Y})\right)^2\right) + 2\left(E(\hat{Y}) - \hat{Y}\right)\left(E(\hat{Y}) - E(\hat{Y})\right) + \left(E(\hat{Y}) - \hat{Y}\right)^2$$

$$ECMP(\hat{Y}, Y) = E\left(\left(\hat{Y} - E(\hat{Y})\right)^2\right) + \left(E(\hat{Y}) - \hat{Y}\right)^2$$

$$ECMP(\hat{Y}, Y) = V(\hat{Y}) + \left(\text{Sesgo}(\hat{Y})\right)^2$$

Anexo 2. Verosimilitud marginal de la data

La verosimilitud marginal se define por:

$$p(Y|\lambda, p, q, d) = \iint p(Y|\beta, \sigma_\varepsilon^2) p(\beta, \sigma_\varepsilon^2|\lambda, p, q, d) d\beta d\sigma_\varepsilon^2$$

Para obtenerla, (en el caso FB-MIDAS) se puede partir de las funciones de verosimilitud y de la distribución *a priori* presentadas, de la siguiente manera (Akaike 1979a):

$$p(Y|\lambda, p, q, d) = p(Y|\beta, \sigma_\varepsilon^2) p(\beta, \sigma_\varepsilon^2|\lambda, p, q, d)$$

$$p(Y|\lambda, p, q, d) = \left(\frac{1}{2\pi\sigma_\varepsilon^2} \right)^{\frac{nm+p+q+1}{2}} |\lambda R'(RR')^{-1}R|^{\frac{1}{2}} \dots$$

$$\left(\frac{1}{\sigma_\varepsilon^2} \right) e^{-\frac{1}{2\sigma_\varepsilon^2}(\beta - b_{sg})'(X'X + \lambda R'(RR')^{-1}R)(\beta - b_{sg})} e^{-\frac{1}{2\sigma_\varepsilon^2}(|Y|^2 - b_{sg}'(X'X + \lambda R'(RR')^{-1}R)b_{sg})}$$

$$p(Y|\lambda, p, q, d) = \left(\frac{1}{2} \right) \left(\frac{1}{\pi} \right)^{\frac{nm}{2}} \Gamma\left(\frac{nm}{2}\right) (|Y|^2 - b_{sg}'(X'X + \lambda R'(RR')^{-1}R)b_{sg})^{-\frac{nm}{2}} \dots$$

$$|X'X + \lambda R'(RR')^{-1}R|^{-\frac{1}{2}} |\lambda R'(RR')^{-1}R|^{\frac{1}{2}}$$

Al ser un problema no lineal, se puede resolver de la siguiente manera: $-2\ln(p(Y|\lambda, p, q, d))$, ignorar términos constantes y minimizar el resultado. De esta manera:

$$L(\lambda, p, q, d) = nm \ln\left(|Y|^2 - b_{sg}'(X'X + \lambda R'(RR')^{-1}R)b_{sg}\right) + \ln|X'X + \lambda R'(RR')^{-1}R|$$

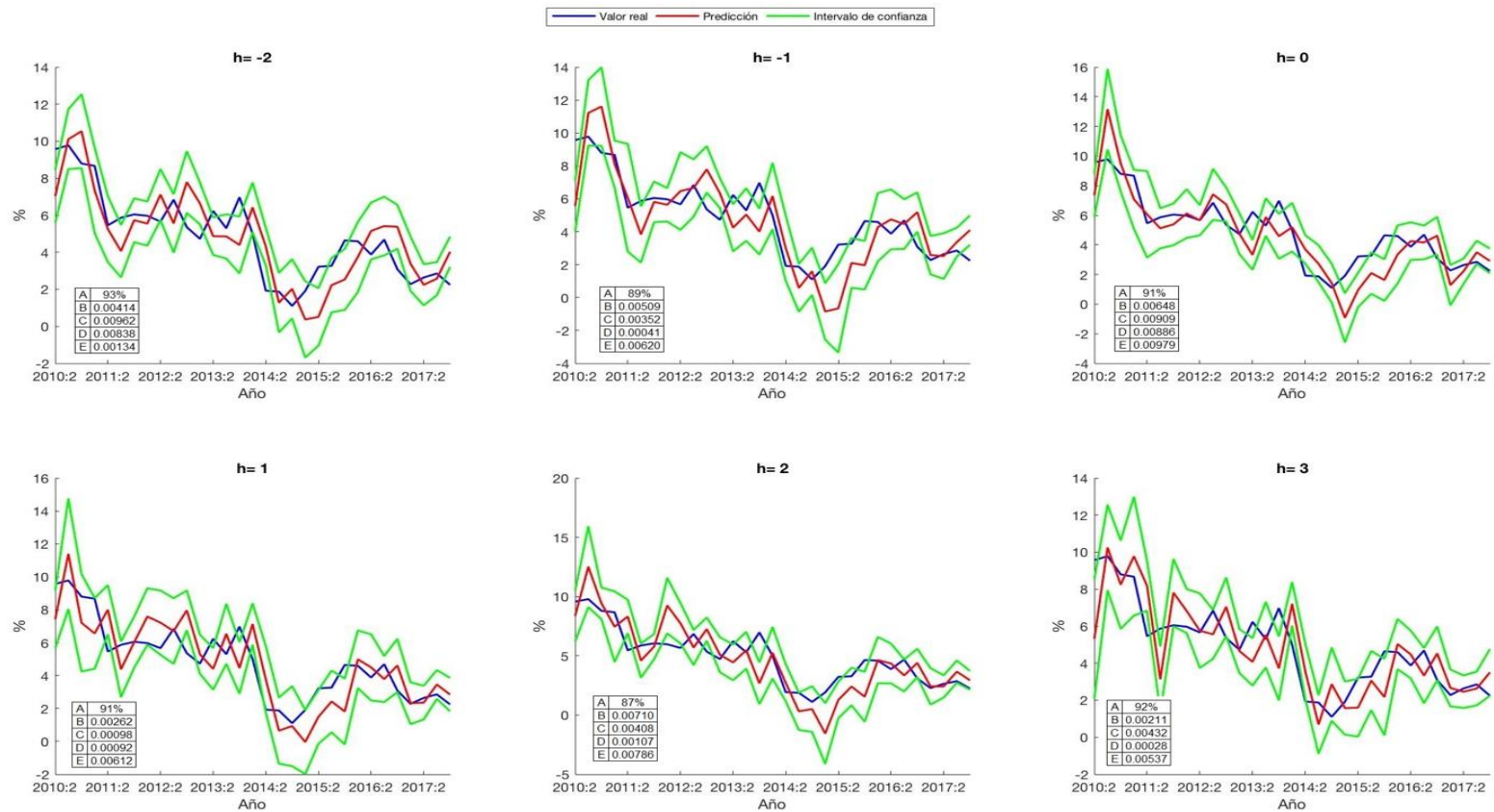
$$- \ln|\lambda R'(RR')^{-1}R|$$

con lo cual, si se tienen \mathcal{R} diferentes modelos ($\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3, \dots, \mathcal{M}_{\mathcal{R}}$)⁴⁰, se escogerá aquel con el menor $L(\lambda, p, q, d)$ (Akaike 1979b).

⁴⁰ Notar que se trata de un espacio discreto.

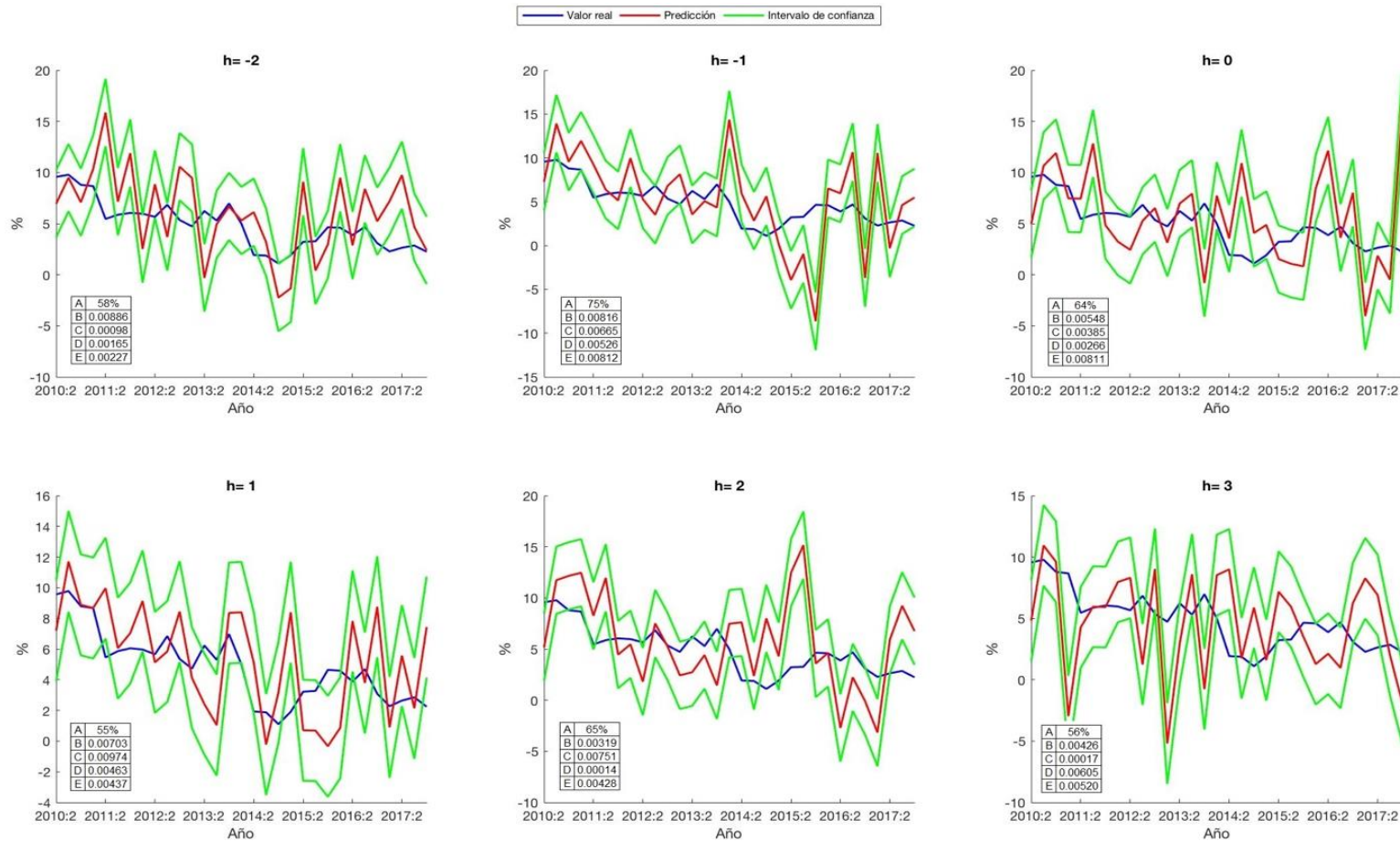
Anexo 3. Predicciones de data empírica con la metodología FB-MIDAS bajo diferentes disparidades

PBI real trimestral vs. indicadores líderes mensuales



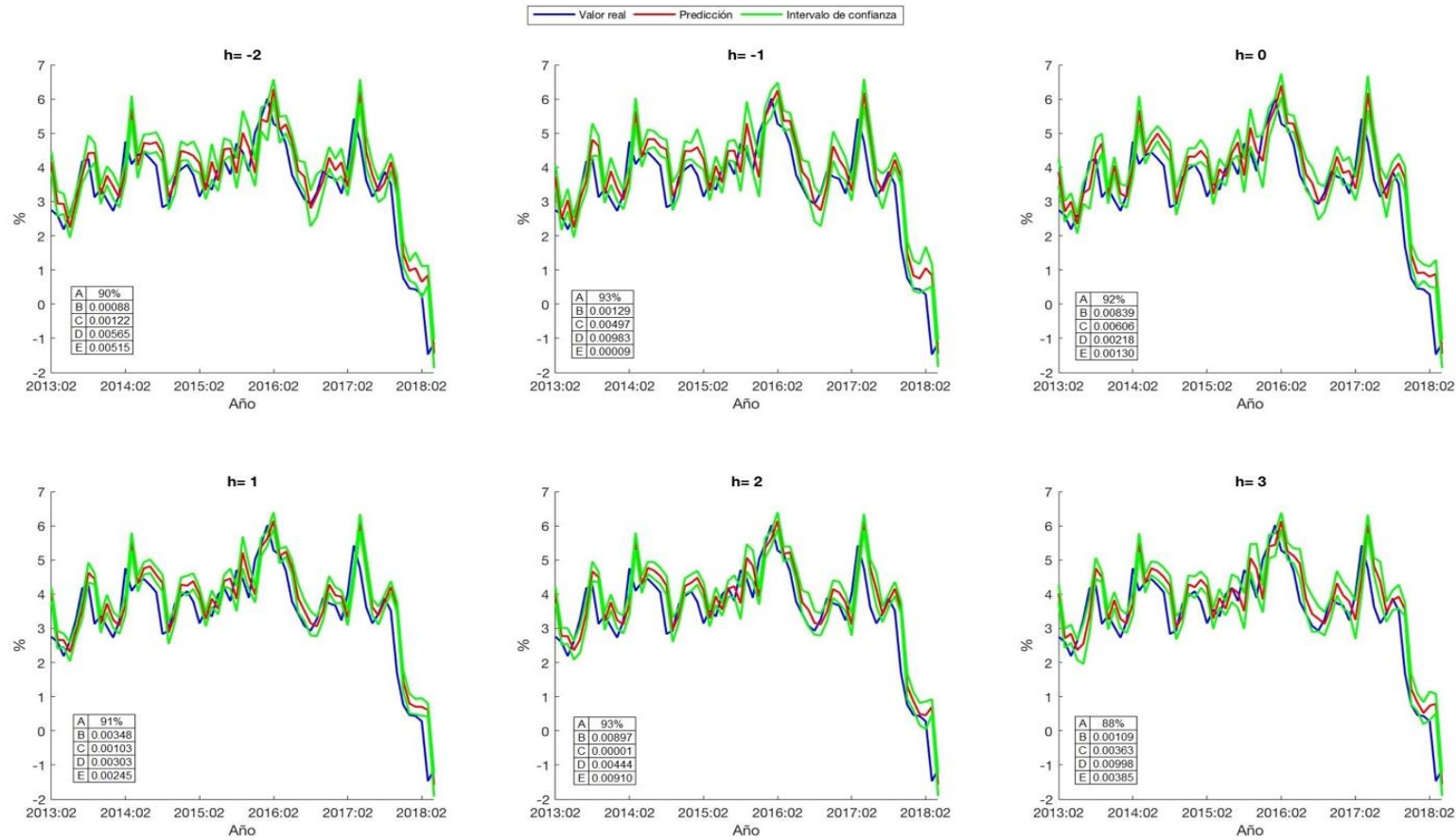
El gráfico muestra la predicción de PBI real trimestral usando indicadores líderes mensuales. Asume una especificación no condicional, una elección BMA y ningún rezago en publicación. Adicionalmente, para cada caso se muestra: (A) la tasa de cobertura de la estimación FB-MIDAS, (B) el p-valor de la prueba Diebold y Mariano (DM) para FB-MIDAS vs. MIDAS, (C) el p-valor de la prueba Diebold y Mariano (DM) para FB-MIDAS vs. U-MIDAS, (D) el p-valor de la prueba Harvey, Leybourne y Newbold (HLN) para FB-MIDAS vs. MIDAS y, (E) el p-valor de la prueba Harvey, Leybourne y Newbold (HLN) para FB-MIDAS vs. MIDAS. Elaboración propia a través de Matlab ®.

PBI real trimestral vs. indicadores líderes diarios



El gráfico muestra la predicción de PBI real trimestral usando indicadores líderes diarios. Asume una especificación condicional, una elección BMS y un rezago en publicación. Adicionalmente, para cada caso se muestra: (A) la tasa de cobertura de la estimación FB-MIDAS, (B) el p-valor de la prueba Diebold y Mariano (DM) para FB-MIDAS vs. MIDAS, (C) el p-valor de la prueba Diebold y Mariano (DM) para FB-MIDAS vs. U-MIDAS, (D) el p-valor de la prueba Harvey, Leybourne y Newbold (HLN) para FB-MIDAS vs. MIDAS y, (E) el p-valor de la prueba Harvey, Leybourne y Newbold (HLN) para FB-MIDAS vs. MIDAS. Elaboración propia a través de Matlab ®.

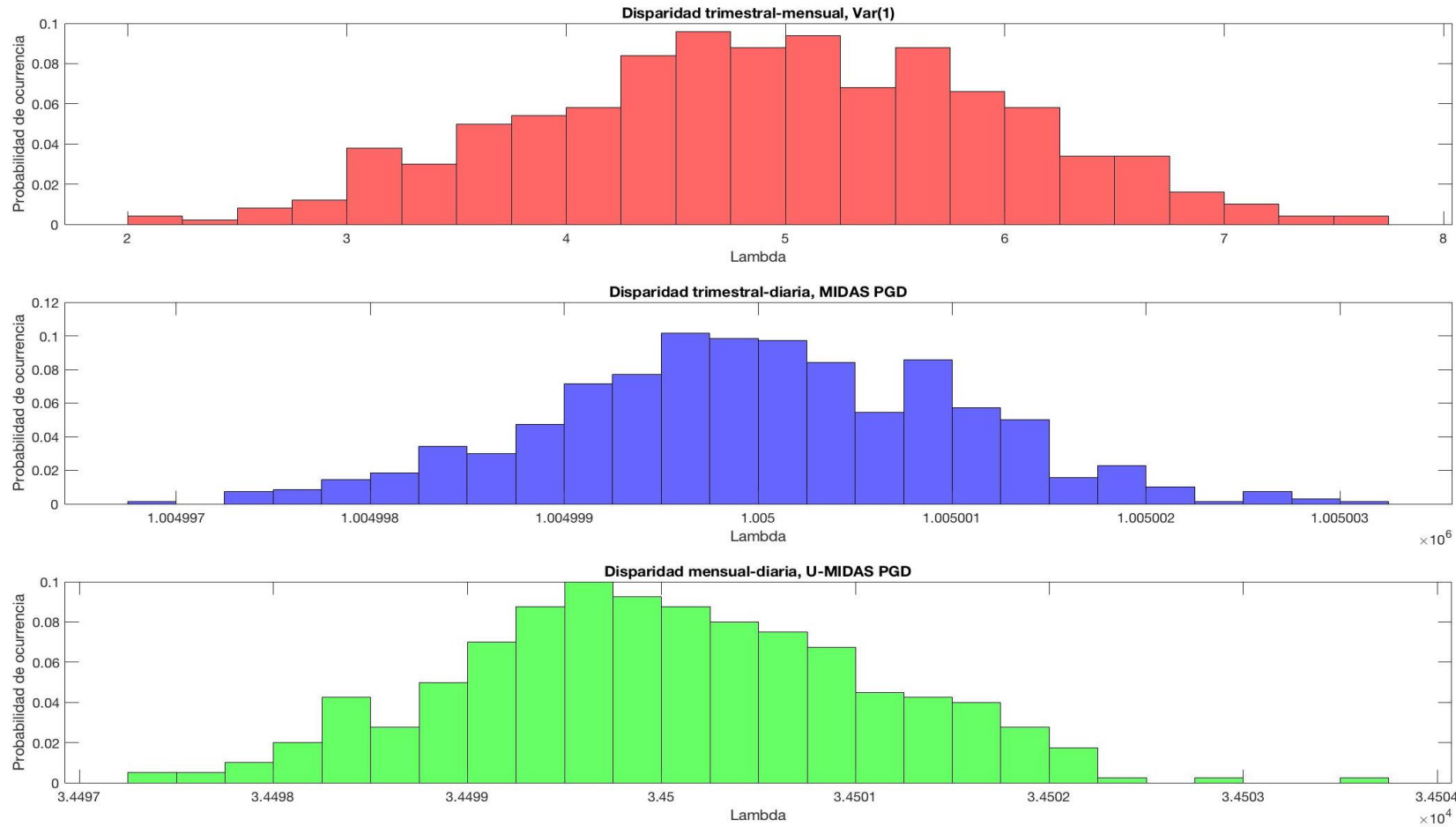
Inflación de alimentos mensual vs. indicadores líderes diarios



El gráfico muestra la predicción de inflación de alimentos usando indicadores líderes diarios. Asume una especificación no condicional, una elección BMS y ningún rezago en publicación. Adicionalmente, para cada caso se muestra: (A) la tasa de cobertura de la estimación FB-MIDAS, (B) el p-valor de la prueba Diebold y Mariano (DM) para FB-MIDAS vs. MIDAS, (C) el p-valor de la prueba Diebold y Mariano (DM) para FB-MIDAS vs. U-MIDAS, (D) el p-valor de la prueba Harvey, Leybourne y Newbold (HLN) para FB-MIDAS vs. MIDAS y, (E) el p-valor de la prueba Harvey, Leybourne y Newbold (HLN) para FB-MIDAS vs. MIDAS. Elaboración propia a través de Matlab ®.

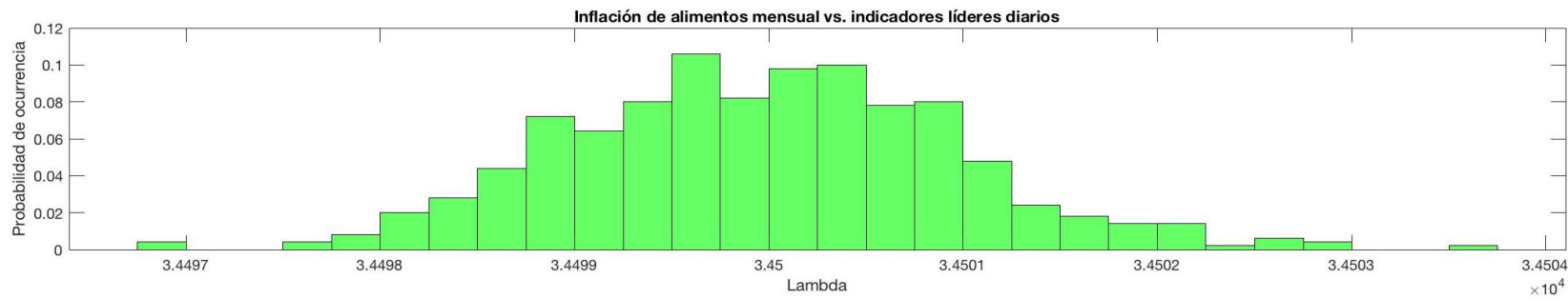
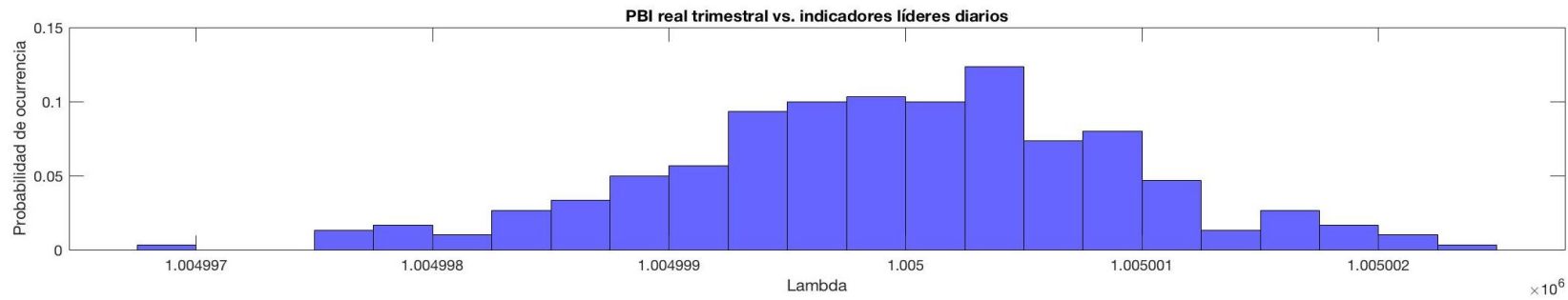
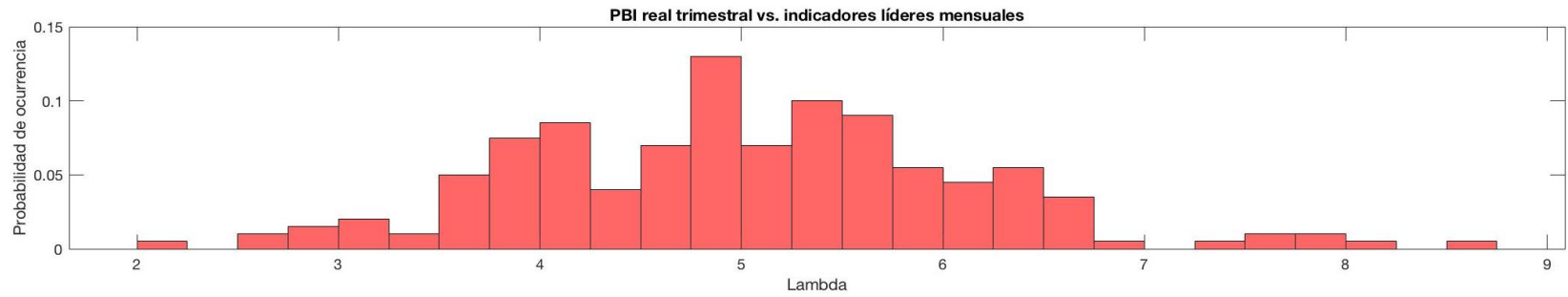
Anexo 4. Valores de lambda óptimo por BMS

Data simulada



Elaboración propia a través de Matlab ®.

Data empírica



Elaboración propia a través de Matlab ®.

Nota biográfica

Renzo Enrique Miní Cuadros

Nació en Lima, el 13 de enero de 1994. Economista, bachiller de la Universidad del Pacífico. Cuenta con las certificaciones de finanzas CFA II, CAIA I y FRM I; y con las de idiomas FCE en inglés y B2 en francés.

Tiene un año y medio de experiencia en estrategia de inversiones y productos de inversión en Credicorp Capital Servicios Financieros S.A.