



**PRONÓSTICO DE LA INFLACIÓN CON MODELOS MIDAS: EVIDENCIA PARA
COLOMBIA**

ISAAC HURTADO RIVERA

TESIS

PAULA MARIA ALMONACID HURTADO

UNIVERSIDAD EAFIT

ESCUELA DE CIENCIAS APLICADAS E INGENIERÍA

MAESTRÍA EN CIENCIAS DE LOS DATOS Y LA ANALÍTICA

MEDELLÍN

2026

Pronóstico de la inflación con modelos MIDAS: evidencia para Colombia

Isaac Hurtado Rivera

Resumen

La inclusión de variables muestreadas en diferentes frecuencias es un desafío empírico en la ciencia económica. Mientras las series macroeconómicas son publicadas mensual o trimestralmente, las series financieras están disponibles diariamente. Una práctica común es agregar o promediar las variables de mayor frecuencia como, por ejemplo, las mensuales o diarias para poderlas incorporar en un mismo modelo. Sin embargo, esto último puede llegar a descartar información y alterar la dinámica temporal entre variables. La regresión MIDAS (Mixed Data Sampling) ofrece una solución a este problema, que además controla la proliferación de parámetros y puede producir estimadores insesgados y eficientes. En un ejercicio aplicado a la inflación mensual en Colombia, se evalúa empíricamente si la información de alta frecuencia mejora el desempeño de pronóstico y si se justifica el uso de modelos tipo MIDAS. Los resultados muestran que la inflación mensual esta parece ser un proceso poco persistente que puede ser modelado con un MIDAS restringido, mientras que el U-MIDAS tiende a caer en sobreajustes. Además, se obtienen predicciones más precisas y menos volátiles cuando se combinan pronósticos.

Palabras clave: Modelos MIDAS, inflación, combinación de pronósticos, NLS, Factores dinámicos, Colombia.

Código JEL: C22, C53

Abstract

Including variables sampled at different frequencies is an empirical challenge in economics. While macroeconomic series are typically released monthly or quarterly, financial series are available daily. A common practice is to aggregate or average the higher-frequency variables (e.g., monthly or daily data) in order to incorporate them into a single model. However, doing so can discard information and distort the temporal dynamics across variables. MIDAS (Mixed Data Sampling) regressions provide a solution to this problem, while also controlling parameter proliferation and yielding unbiased and efficient estimators. Using an application to monthly inflation in Colombia, this study empirically assesses whether high-frequency information improves forecasting performance and whether the use of MIDAS-type models is warranted. The results suggest that monthly inflation is a low-persistence process that can be adequately modeled with a restricted MIDAS specification, whereas U-MIDAS tends to overfit. In addition, forecasts become more accurate and less volatile when combined forecast methods are used.

Key words: MIDAS models, inflation, forecast combination, NLS, dynamic factors, Colombia.

1. Introducción

Una de esas variables económicas que más atención recibe alrededor del mundo es la inflación. Las empresas la consideran en su presupuesto y/o la tienen en cuenta para evaluar la rentabilidad de su negocio o de los proyectos de inversión, los hacedores de políticas la siguen de cerca para poder actuar en consecuencia con sus instrumentos de, por ejemplo, política monetaria; en el caso de los mercados financieros determina la reasignación eficiente de los recursos entre diferentes tipos activos y la recomposición de los portafolios de inversión para obtener el mayor retorno posible y, en el caso de la academia, es objeto de estudio para poder entender sus determinantes, causas, consecuencias y comportamiento.

Un problema muy común al estudiar y/o modelar diferentes variables tanto en la economía como en las finanzas es que las variables se muestrean en diferentes frecuencias. Por ejemplo, los datos agregados de cuentas nacionales como el PIB, la inversión o el consumo son publicados de forma trimestral, mientras que los resultados de mercado laboral o la evolución de los precios suelen estar disponibles mensualmente, mientras las tasas de interés y series financieras están disponibles de forma diaria o, en el peor de los casos, semanal. Esto plantea un desafío: ¿cómo modelar la relación entre una serie de baja frecuencia, por ejemplo, trimestral o mensual, y predictores de alta frecuencia (mensual, semanal, diaria) usando de la mejor manera toda la información disponible?

Tradicionalmente, se es agregan o promedian las variables de alta frecuencia para convertirlas en series de baja frecuencia de forma que todas las variables queden alineadas y sea posible estimar el modelo. Sin embargo, varios artículos resaltan que esta práctica tan común puede conllevar a tener estimadores sesgados y menos eficientes lo que puede conducir a conclusiones erradas; además, se pierde información relevante de las series de alta frecuencia que pueden ayudar a explicar la serie de baja frecuencia (Armesto, et al. 2010; Andreou, et al. 2010a; Ghysels, et al. 2004a). Por ejemplo, para hacer una regresión que impliquen información financiera como la tasa de cambio y su impacto en las exportaciones netas de una economía, la diferencias de frecuencias hace que se tome el promedio de la tasa de cambio en el mes o en el trimestre para que tenga la misma frecuencia que las exportaciones y las importaciones, ignorando diferentes choques o innovaciones que pudieron ocurrir dentro ese periodo de tiempo y que pueden tener efectos asimétricos sobre la balanza comercial.

Es por esto que han surgido diversas metodologías orientadas a incorporar toda la información disponible en diferentes frecuencias en un mismo modelo. Ghysels, et al. (2004a) proponen la metodología MIDAS (Mixed Data Sampling) que permite abordar este tipo de problemas y generar estimaciones consistentes y eficientes. Además, la evidencia

muestra que se presentan ganancias en la capacidad predictiva en comparación con las practicas tradicionales de agregación o modelos puente.

Posterior al trabajo de Ghysels, et al. (2004a) se ha abierto un abanico de posibilidades en las que se aplica el MIDAS para el pronóstico de diferentes variables macroeconómicas que han mostrado tener gran poder predictivo como, por ejemplo, AR-MIDAS (Clements & Galvão, 2008) que incluye un componente autorregresivo, el Unrestricted-MIDAS (U-MIDAS) (Forni, et al. 2012) que sería una versión más flexible del MIDAS e, incluso, el uso de factores dinámicos junto con el esquema MIDAS o Factor-MIDAS (Marcellino & Schumacher, 2010; Galeano, et al. 2021).

En este trabajo se estiman y comparan diferentes especificaciones de modelos tipo MIDAS con el fin de (i) caracterizar propiedades empíricas de la serie de la inflación mensual, (ii) evaluar si explotar información de alta frecuencia mejora el desempeño de las predicciones del modelo y (iii) construir un pronóstico robusto para Colombia apalancado en la combinación de estos. Para ello, se identifican las variables que, según la teoría económica y la evidencia empírica, tienen poder explicativo sobre la inflación para luego estimar diferentes especificaciones dada una grilla de posibles parámetros y rezagos. Después, se evalúa si la especificación es adecuada, se comparan los modelos y, usando diferentes formas de combinación de pronósticos, se encuentra la mejor estimación de inflación posible.

En concreto, se utilizó un modelo de factor MIDAS que permite reducir la dimensionalidad de la información empleada en la predicción al tiempo que captura relaciones complejas entre las variables y se reduce la cantidad de parámetros a estimar, aunque se pierde interpretabilidad de los parámetros. Una vez obtenidos estos factores se estimaron diferentes especificaciones de modelo MIDAS, se testeó si las restricciones impuestas son o no pertinentes de forma que se pueda comparar entre MIDAS y U-MIDAS. Por último, se evaluaron métricas in-sample y out-of-sample de cada modelo que ayudan a inferir algunas propiedades del proceso generador de la inflación y si este tipo de modelos son los adecuados para modelar esta variable. Por último, se combinaron pronósticos de diferentes formas para tener una estimación robusta de la inflación mensual.

Los resultados muestran que, (i) los modelos MIDAS tienen una capacidad de predicción superior a la de modelos de referencia como los modelos autorregresivos de orden 1 (AR(1)), (ii) la especificación U-MIDAS tiene mejores métricas in-sample de error frente al MIDAS tradicional o restringido pero peores out-of-sample lo que sugiere posibles problemas de sobreajuste y (iii) la combinación de pronósticos genera predicciones más precisas y menos volátiles frente a solo considerar una sola especificación reduciendo la incertidumbre del modelo.

A pesar de que el modelo MIDAS no es una metodología nueva, al igual que la combinación de pronósticos, su uso sigue siendo escaso en ejercicios aplicados para Colombia fuera de los

ejercicios para el PIB de Galeano, et al. (2021) y Cardenas, et al. (2020) para la inflación de alimentos y de González (2010) en el caso de la combinación de pronósticos. En este contexto, el aporte de este artículo es nutrir la evidencia empírica en Colombia sobre las ventajas del uso del MIDAS en la predicción, en este caso, de la inflación total no solo en la de los alimentos como lo hace Cardenas, et al. (2020); y evaluar los beneficios de tener combinaciones de pronósticos.

El trabajo se organiza de la siguiente manera. La sección 2 presenta una revisión de literatura sobre los modelos de la familia MIDAS, su origen, algunas diferencias entre ellos o con modelos de otras características y algunos casos aplicados en la combinación de pronósticos, luego en la sección 3 se presenta de formalización metodológica de estos modelos. Finalmente, la sección 4 pasar a presentar los resultados de la estimación y las conclusiones.

2. Revisión de literatura

Esta sección presentará el origen de los modelos MIDAS y sus diferentes variantes, los beneficios y posibles deficiencias frente a modelos puente y modelos alternativos que pueden tratar este mismo problema como el filtro de Kalman; y el uso de la combinación de pronósticos unido al uso de la regresión MIDAS.

Origen

Ghysels, et al. (2004a) estudian el modelo ICAPM o CAPM (Capital asset pricing model) intertemporal de Merton (1973) que describe lo que los autores llaman la primera ley fundamental de las finanzas: A mayor riesgo (varianza), mayor retorno. Aunque parezca una obviedad actualmente, este trade-off había sido esquivo en la evidencia empírica ya que los artículos que intentaban estimar esta relación encontraban relaciones negativas, estadísticamente no significativas o eran hallazgos no robustos.

El principal problema a la hora de hacer esta estimación es encontrar un buen estimador para la varianza condicionada ya que esta última no es observable y debe ser estimada o filtrada a partir de los retornos pasados. Para abordar este problema empírico, los autores proponen el estimador MIDAS que permite tener datos sobre el retorno de los mercados en una frecuencia mayor, usando una forma funcional parsimoniosa y flexible que pondera los retornos diarios pasados al cuadrado para estimar la varianza condicionada. Este estimador es contrastado con los tradicionales modelos de la familia GARCH (*Generalized Autoregressive Conditional Heteroscedasticity*) y el Rolling Window de French et al. (1987).

Los autores encuentran que hay una relación positiva entre el riesgo y el retorno en el mercado accionario, conclusión que es robusta ante diferentes muestreos, cambios en las funciones ponderadoras, a la incorporación de asimetrías y a la inclusión de variables relacionadas con el ciclo económico. Entre las ventajas que se tienen al usar MIDAS frente

a los otros estimadores, es que, al considerar mucha más historia de los retornos realizados (252 días un año bursátil) se obtiene una mayor persistencia de los efectos de la varianza condicionada en la media condicional. Esto último, es especialmente importante cuando se consideran asimetrías pues encuentran que, al igual que en los GARCH, existen efectos heterogéneos de las noticias negativas en los mercados, pero la persistencia de las noticias positivas al parecer son mucho más altas que las negativas que no tienen efectos más allá de un mes, cuestión que no es capturada en los GARCH

Los hallazgos de la evidencia empírica hasta ese momento habían sido de alguna manera, sesgados por la poca historia que se tomaba en cuenta a la hora de analizar la relación riesgo-retorno. De hecho, cuando amplían la ventana de la Rolling Window a más de un mes el parámetro empieza a crecer y se vuelve estadísticamente significativo. Además, remarcan que, el estimador MIDAS tiene mejores propiedades que el GARCH y el Rolling Windows ya que es insesgado tanto en regímenes de baja y alta volatilidad por lo que logra capturar mejor la varianza realizada para el periodo analizado frente a, por ejemplo, el GARCH que tiende a suavizar la media del proceso lo que sugiere un sesgo.

Modelos puente vs MIDAS

Andreou, et al. (2010a) presentan las propiedades asintóticas y las implicaciones empíricas del estimador MIDAS comparados con de los modelos tradicionales que agregan o promedian los datos para estimar un modelo en la misma frecuencia. Los autores remarcan que un esquema que considere una ponderación decreciente como la que plantea MIDAS con las funciones Beta o Almon, debería ser más apropiada que una que asuma pesos iguales o flat aggregation ya que llevaría a estimadores ineficientes y/o sesgados, ya que se ignoran las dinámicas temporales de las series subyacentes. En concreto, dicen que la diferencia en eficiencia de los estimadores radica en que, si una serie de alta frecuencia es persistente o tiene estructuras heterocedásticas, promediar lleva a desperdiciar información y a generar mayor error. Así, entre mayor es la diferencia entre las series de baja y alta frecuencia mayor error (menor eficiencia) habrá.

En un ejercicio empírico muestran como la regresión MIDAS puede explicar mejor las diferencias de crecimiento económico mediante el modelo de Solow de Mankiw, et al. (1992). Los resultados, en general, desde el punto de vista intuitivo son similares a Mankiw, et al. (1992) con respecto a la convergencia condicionada, sin embargo, encuentran (i) que las estimaciones con MIDAS son más eficientes y (ii) que los parámetros del capital físico son mayores, mientras que los de capital humano son menores, consistente con sus simulaciones sobre el sesgo de los parámetros del flat aggregation.

El trabajo de Ghysels, et al. (2004b) se enfoca en explicar la naturaleza del sesgo y la pérdida de eficiencia entre los modelos tipo MIDAS con los de rezagos distribuidos (ADL) que

agregan o promedian series de alta frecuencia. El artículo examina dos problemas que aparecen cuando se trata de modelar datos de distinta frecuencia: el sesgo por agregación y las distorsiones espectrales o aliasing. El primero se refiere al sesgo que ocurre cuando se promedia o se suman datos de alta frecuencia para bajarle la frecuencia, alterando la relación verdadera entre las variables; mientras el segundo se refiere a la distorsión que surgen cuando un proceso continuo se observa de forma discreta cada cierto tiempo.

Dentro del sesgo por agregación, muestran que tipo de muestreo de las variables puede llevar a un sesgo adicional en los parámetros. En concreto, si la variable dependiente e independiente se miden con distintos esquemas de muestreo, es decir, si una es flujo y la otra es un valor puntual, como puede ser el caso de una regresión del consumo y la tasa de interés, se debe buscar la manera de agregar o promediar la variable de valor puntual o point sampling, que es la cuestión central de la metodología. En el caso donde ambas variables sean point sampling o flujo no hay sesgo adicional. Además, Ghysels, et al. (2004b) muestran que el sesgo por aliasing, se reduce al aumentar la frecuencia de muestreo de la variable independiente. Los autores demuestran que cuando variable dependiente y las independientes se observan en la misma frecuencia, las estimaciones del ADL y el MIDAS son equivalentes y consistentes. No obstante, si las variables explicativas están a una mayor frecuencia, como ocurre en la realidad, las estimaciones del ADL son menos eficientes y potencialmente sesgadas frente al MIDAS que controla mejor el aliasing y usa más información del proceso verdadero de alta frecuencia dada su función ponderadora.

En síntesis, argumentan que el MIDAS es una generalización de los modelos de rezagos distribuidos, que, en lugar de promediar o agregar, introduce una función de ponderación parsimoniosa se logran tener estimadores consistentes y más eficientes, ya que se (i) logra aprovechar toda la información disponible de la serie de alta frecuencia, (ii) no da el mismo peso a todas las observaciones sino que considera que los datos más recientes son más informativos, (iii) no se distorsiona la verdadera relación entre las variables ya que captura mejor la dinámica de las series y (iv) no se cae en sobreparametrización.

MIDAS vs AR-MIDAS

En los artículos de Clements y Galvão (2008) y Clements y Galvão (2009) se analizan (i) el impacto de datos en tiempo real frente a datos revisados en el error de predicción, (ii) si es preferible combinar pronósticos de modelos con un solo indicador adelantado o, por el contrario, incorporar varias de estas variables en un mismo modelo y, además, (iii) proponen una extensión del modelo MIDAS al incorporar un término autorregresivo de la variable dependiente o un AR-MIDAS. Los autores muestran que la especificación del AR-MIDAS debe ser ajustada, pues, de lo contrario, se le podría generar un componente estacional a las variables de alta frecuencia, incluso cuando estos no presentan estacionalidad.

En términos de resultados encuentran que (i) usar la información en tiempo real da ganancias en la predicción frente al uso de datos revisados para modelos que usen indicadores líderes, especialmente a horizontes cortos de tiempo, (ii) el modelo MIDAS tiene menores errores de pronósticos frente a modelos AR y ADL en el corto plazo, (iii) los pesos que le da la función ponderadora de MIDAS a las observaciones de alta frecuencia son diferente conforme cambia el horizonte de predicción y las variables, (iv) la inclusión de un componente autorregresivo mejora la capacidad predictiva de los modelos y (v) hay evidencia de los beneficios de combinar pronósticos.

Midas vs U-MIDAS

Otro desafío que surge a la hora de mezclar frecuencias en los datos, aparte de los mencionados por Ghysels, et al. (2004b) es la proliferación de parámetros. A pesar de que la agregación plana como se acaba de mostrar puede llevar a parámetros sesgados e ineficientes junto con la pérdida de información y de la verdadera relación entre las variables; es una solución parsimoniosa (Armesto, et al. 2010), aunque no la mejor, de afrontar este obstáculo.

Para evitar este aumento de los parámetros la regresión MIDAS propuesta por Ghysels, et al. (2004a) impone una función ponderadora o de agregación con decaimiento, de forma que se reducen los parámetros a estimar. Sin embargo, la inclusión de esta restricción en la ponderación de los parámetros puede llegar a introducir sesgo si la verdadera relación no sigue la forma funcional impuesta y aumenta la complejidad a la hora de la estimación dada la naturaleza no lineal de la función que se le impone a la dinámica de los datos.

Es por esto que surge el MIDAS no restringido (U-MIDAS), que, como su nombre lo dice, no impone restricciones sobre la dinámica de los datos. Así, cada coeficiente es estimado por separado mediante mínimos cuadrados ordinarios (MCO) (Feroni, et al. 2012) lo que simplifica la implementación y reduce el costo computacional. Esta mayor flexibilidad permite generar formas funcionales suaves o simples y tanto similares como diferentes a los del MIDAS de Ghysels, et al. (2004a) lo que permite aproximar la verdadera relación entre los datos. No obstante, esta flexibilidad también conlleva a aumentar la cantidad de parámetros a estimar, en especial, si la diferencia en la frecuencia entre las series es grande, por ejemplo, anual y diaria (Armesto, et al. 2010; Feroni, et al. 2012), lo que aumenta el riesgo de sobreajuste y puede llevar a que los coeficientes pueden salir con cualquier signo y magnitud contraintuitivos, en especial, si la muestra es pequeña

El artículo de Feroni, et al. (2012) derivan la estructura del U-MIDAS, muestran las condiciones sobre las cuales es identificable y comparan la capacidad de predicción del PIB

para Estados Unidos y Europa de esta especificación con el MIDAS clásico de Ghysels, et al. (2004a) o restringido como lo llaman. Entre otras cosas, encuentran que las predicciones tanto del MIDAS como el U-MIDAS superan al modelo de referencia AR(1). Cuando la diferencia en las frecuencias de las series es pequeña y cuando los procesos son más persistentes, el U-MIDAS tiende a capturar mejor la dinámica real de los datos y a mejorar las predicciones comparado con el MIDAS restringido. En contraste, si la diferencia entre series es grande y son menos persistentes el MIDAS restringido presenta mejores pronósticos que el U-MIDAS además de ser más parsimonioso. Intuitivamente, el tener poca diferencia de frecuencias da mayor flexibilidad para estimar los parámetros por lo que se tiene un efecto más “limpio” de cada observación, pero a mayor diferencia el incremento de los parámetros a estimar empeora la precisión predictiva.

En el caso de los pronósticos para Estados Unidos, no hay un ganador estricto entre el U-MIDAS y el MIDAS, cuestión que depende de las variables explicativas y del horizonte de predicción. Mientras in-sample el U-MIDAS parece tener mejores resultados, out-of-sample no hay un dominante claro. Por su parte para el área Euro, el U-MIDAS si es superior al MIDAS, aunque si se considera el periodo de la crisis de 2008 el MIDAS tiene mejor comportamiento. En términos generales, los autores concluyen que el U-MIDAS es un gran competidor, bajo ciertas condiciones, del MIDAS restringido por lo que debe ser considerado como una alternativa a la hora de pronosticar con diferentes frecuencias.

En resumen, se tiene que la agregación plana es parsimoniosa, fácil de estimar pero es sesgada, ineficiente y descarta información sobre las innovaciones; MIDAS es parsimonioso, genera parámetros consistentes y eficientes pero es más difícil de estimar dada la naturaleza no lineal de la forma funcional que se le impone a la dinámica de los datos y el U-MIDAS conserva el comportamiento de las variables, es fácil de estimar por MCO pero es propenso a la proliferación de parámetros si la diferencia de frecuencias es muy grande.

MIDAS vs Filtro de Kalman

En el trabajo de Bai, et al. (2010) analizan las semejanzas y diferencias entre la regresión MIDAS y el filtro de Kalman aplicado a datos de distintas frecuencias. Desde el punto de vista del set-up del modelo, una de las principales diferencias que se marcan es que el filtro de Kalman puede llegar a ser bastante complejo ya que se deben especificar explícitamente un modelo dinámico para todas las series involucradas: las variables y factores latentes de alta frecuencia y las series de baja frecuencia. Este sistema de ecuaciones suele involucrar muchos parámetros, lo que aumenta en el costo computacional y la vulnerabilidad a errores de especificación (Ghysels, et al. 2016).

Bai, et al. (2010) muestran las condiciones bajo las cuales ambos modelos son equivalentes en términos de pronóstico. Bajo los supuestos de que el modelo espacio-estado subyacente

al filtro de Kalman está bien especificado y que no existe incertidumbre sobre los parámetros, es decir, que se conocen los parámetros poblacionales, la estimación por el filtro de Kalman coinciden con las que se obtendría a partir del modelo MIDAS. No obstante, cuando estos supuestos se relajan, por ejemplo, si el modelo en espacio de estados está mal especificado o el número de parámetros es elevado en relación con el tamaño muestral, el MIDAS tiende a ser más robusto, precisamente porque no requiere especificar el proceso generador de datos de todas las variables involucradas. En los ejercicios de simulación del artículo, los autores encuentran que, en muchos escenarios, los errores de pronóstico de MIDAS son muy similares a los del filtro de Kalman, y en algunos casos son menores en horizontes cortos, mientras que en horizontes más largos el filtro de Kalman puede recuperar una ligera ventaja gracias a su naturaleza más estructural.

En cuanto al esquema de agregación, Bai, et al. (2010) demuestran teóricamente que los pronósticos del filtro de Kalman se pueden escribir como un promedio ponderado de los rezagos de las variables de alta frecuencia y que estos pesos pueden ser similares a los obtenidos con el modelo MIDAS con sus funciones de agregación. Al contrastar empíricamente, encuentran que efectivamente no hay diferencias significativas entre los pesos estimados por los dos modelos y que los errores de predicción resultantes son prácticamente iguales.

Por su parte, Andreou, et al. (2010b) marcan otras desventajas del filtro de Kalman frente al uso de la regresión MIDAS. Recalcan que el filtro de Kalman es viable cuando hay pocas variables en consideración, dado que, como mostró Bai, et al. (2010), los modelos espacio-estado requieren una ecuación por cada variable considerada por lo que el número de parámetros a estimar para la ecuación de medición, dinámica de estados y sus procesos de error puede ser grande. En ese sentido, el uso de la regresión de MIDAS es superior en estos casos dada su parsimonia cuando hay muchas variables de alta frecuencia como las financieras y, en especial, cuando hay diferentes frecuencias en un mismo modelo ya que es más fácil caer en malas especificaciones (Ghysels, et al. 2016)). Sin embargo, rescatan el filtro de Kalman que tiene ventajas en la interpretabilidad de los resultados frente al MIDAS y en mirar cómo los diferentes choques afectan los pronósticos.

Combinación de pronósticos y MIDAS

En Andreou, et al. (2010b) y Torres (2015) se presentan ejercicios de predicción del PIB usando MIDAS junto con combinación de pronósticos. Andreou, et al. (2010b) presentan ejercicios para pronosticar la actividad económica real trimestral utilizando datos financieros diarios. Su trabajo se basa en la naturaleza prospectiva de los mercados financieros y en el supuesto de eficiencia de mercado, que consideran que toda la información sobre los riesgos y el estado actual y futuro de la economía es descontada rápidamente y reflejada en los diferentes indicadores e instrumentos financieros del mercado de valores, por lo que es

sumamente valioso e importante considerar datos de alta relevancia a la hora de hacer proyecciones macroeconómicas.

El trabajo presenta una extensión de la metodología MIDAS al aplicarla a los modelos de factores dinámicos como una forma de resumir, rescatar y encontrar los patrones complejos más relevantes para las 69 series macroeconómicas trimestrales y las 991 series financieras diarias que se consideraron. De forma similar a Ghysels, et al. (2004b), el artículo compara la capacidad de predicción de los modelos tradicionales de rezagos distribuidos (ADL) y rezagos distribuidos con factores dinámicos (FADL) de ponderación plana con los MIDAS y MIDAS con factores dinámicos o Factor-MIDAS, usando métodos de combinación de pronósticos con el fin de proporcionar predicciones robustas y precisas de la actividad económica.

Para esto, proponen el MIDAS con adelantos que permite aprovechar la información en tiempo real de las variables financieras para hacer predicciones intraperiodo, así como para otros periodos en el futuro (Armesto, 2010). Sobre este punto, los autores resaltan algunas diferencias sobre el ejercicio de nowcasting y el uso del MIDAS con adelantos: (i) el nowcasting se refiere a actualizaciones de las previsiones dentro de un mismo periodo, mientras que el MIDAS con adelantos no solo puede pronosticar en el periodo actual sino también en el futuro; (ii) Dada las diferencias en el muestreo de las series macroeconómicas, mientras el nowcasting modela explícitamente la llegada irregular de los datos económicos, el MIDAS con adelantos, al usar la información financiera, simplemente tiene que actualizar los nuevos datos. Por último, en el nowcasting se pueden tener problemas con la revisión de los datos macroeconómicos, el MIDAS con adelantos al usar información financiera no se está expuesto a, por ejemplo, errores de medición.

Con respecto a las combinaciones de pronósticos, Andreou, et al. (2010b) dicen que puede dar predicciones más precisas ya que, de alguna manera, se atacan los problemas de especificación al considerar muchos modelos al mismo tiempo en lugar de uno solo, reduciendo la incertidumbre de modelo. Así, el tener en cuenta, casi de forma simultánea, diferentes especificaciones de los modelos como la incorporación de diferentes conjuntos de predictores y las diversas estructuras de rezagos en los modelos (i) reducen el sesgo de cada modelo y (ii) dan pronósticos más robustos a periodos de inestabilidad o a choques estructurales, por lo que (iii) tienen mejor comportamiento por fuera de la muestra.

En cuanto a los resultados, usando la combinación de pronósticos encuentran que (i) los datos financieros como los relacionados con el riesgo corporativo, los títulos del gobierno y las acciones ayudan a mejorar los pronósticos trimestrales del crecimiento económico para Estados Unidos comparado con modelos de referencia como los AR, la caminata aleatoria e, incluso, encuestas de pronósticos; (ii) los modelos tipo MIDAS tienen predicciones con menores errores comparado con los modelos tradicionales de agregación plana como los ADL o FADL, (iii) el uso de factores dinámicos trimestrales genera ganancias a la hora de

proporcionar pronósticos frente a modelos que no los incorporan como ADL y MIDAS y (iv) la incorporación de factores diarios contribuye a tener un mejor desempeño en las proyecciones macroeconómicas en contraste con los modelos que solo tienen factores trimestrales. Además, las ventajas del uso de los modelos MIDAS en la predicción son mayores cuando se tienen adelantos en la especificación del modelo.

De forma análoga, Torres (2015) hace un ejercicio similar de predicción del PIB usando los modelos MIDAS usando combinación de pronósticos para algunos países de Europa. Con respecto a la combinación de pronósticos, es una práctica bastante utilizada en la literatura con aplicaciones en crecimiento económico, inflación, tasa de cambio y retornos de activos financieros y ayuda, especialmente, en situaciones donde existan relaciones entre variables muy complejas. Aunque existe un consenso de que sirven para generar predicciones más robustas, la discusión está en la forma de ponderar estos pronósticos.

Mientras Andreou, et al. (2010b) combinan los pronósticos por el error cuadrático medio out-of-sample, Torres (2015) lo hace mediante promedios bayesianos que pondera aquellas especificaciones que tienen mayor probabilidad posterior, priorizando así, aquellas variables que consistentemente son incorporadas en varios modelos y cuyos parámetros estimados tienen una relativa estabilidad respecto a otras variables. En general, también encuentra que al integrar MIDAS con la combinación de pronósticos se genera una mejora en la capacidad predictiva comparada con un AR(1).

3. Metodología

La idea de esta sección es la formalización de la regresión MIDAS y sus variantes. Como se mencionó anteriormente, se estiman diferentes especificaciones del modelo MIDAS que se encuentran en el paquete “midasr” de R (Ghysels, et al. 2016) y seguirá de cerca el trabajo de Andreou, et al. (2010b), ya que se considerará un modelo MIDAS con factores dinámicos junto con la combinación de pronósticos.

. Considerando el siguiente modelo de rezagos distribuidos con un solo regresor por simplicidad y en un entorno donde todas las variables están en la misma frecuencia:

$$Y_t = \beta_0 + \sum_{j=0}^n \beta_j Z^j X_t + e_t \quad (1)$$

Donde β_j son los coeficientes asociados a los n rezagos de baja frecuencia considerados de X_t y Z es el operador de rezagos (de ahora en adelante operador de rezagos de baja frecuencia). Suponiendo ahora que se enfrenta al problema de series en diferentes frecuencias, sea Y_t la variable en baja frecuencia por ejemplo trimestral o mensual y x_t la serie en alta frecuencia como lo podría ser una diaria. Como lo muestran Armesto, et al.

(2010), el método más simple para igualar las frecuencias y estimar el modelo es usar la agregación o el promedio de la serie de alta frecuencia, usando como divisor m la cantidad de veces que se muestrea x_t en una sola muestra de y_t así:

$$\bar{X}_t = \frac{1}{m} \sum_{k=0}^{m-1} L^{\frac{k}{m}} x_t = \frac{x_{t-\frac{0}{m}} + x_{t-\frac{1}{m}} + \dots + x_{t-\frac{m-1}{m}}}{m} \quad (2)$$

Donde $L^{\frac{k}{m}}$ es el operador de rezago para la serie de baja frecuencia de forma que $L^{\frac{k}{m}} \cdot x_t = x_{t-\frac{k}{m}}$ y denota un rezago fraccional de la variable de alta frecuencia dentro del periodo t de la variable de baja frecuencia. Por ejemplo, $x_{t-\frac{0}{m}}$ representa el valor de la variable de alta frecuencia al final del periodo t ; $x_{t-\frac{1}{m}}$ sería el valor de x un día antes de finalizar el mes y así sucesivamente hasta el primer subperiodo del periodo t o lo que sería el primer día del mes en consideración.

Siendo así, sustituyendo (2) en (1) se obtiene el modelo de flat aggregation o modelos puente (Armesto, et al. 2010; Andreou, et al. 2010a):

$$Y_t = \beta_0 + \sum_{j=0}^n \beta_j Z^j \bar{X}_t + e_t \quad (3)$$

Como lo comentan Armesto, et al. (2010) esta forma funcional asume que para cada observación individual de x_t tienen el mismo peso, lo que conlleva a estimaciones sesgadas y poco eficientes. Para lidiar con esto Ghysels, et al. (2004) proponen MIDAS (Mixed data sampling) que modela los coeficientes de las variables explicativas como una función de rezago distribuido de forma parsimoniosa (Clements & Galvão, 2009):

$$Y_t = \beta_0 + \beta_1 \sum_{k=0}^K \Phi(k, \boldsymbol{\theta}) L^{\frac{k}{m}} x_t + e_t \quad (4)$$

Siendo $\Phi(k, \boldsymbol{\theta})$ una función ponderadora parametrizada por el vector de parámetros $\boldsymbol{\theta}$ que determina los pesos de la agregación temporal, n el número de rezagos de baja frecuencia considerados y $K = (m \cdot n) - 1$ es el número total de rezagos de alta frecuencia. Para un modelo que reúna datos mensuales y diarios con 3 meses de datos semanales ($m = 4$) la sumatoria sería:

$$Y_t = \beta_0 + \beta_1 \left[\Phi(0, \boldsymbol{\theta}) x_{t-\frac{0}{4}} + \Phi(1, \boldsymbol{\theta}) x_{t-\frac{1}{4}} + \Phi(2, \boldsymbol{\theta}) x_{t-\frac{2}{4}} + \Phi(3, \boldsymbol{\theta}) x_{t-\frac{3}{4}} + \dots \right] + e_t \quad (5)$$

Ghysels, et al. (2004a), proponen la función beta como especificación para $\Phi(k, \boldsymbol{\theta})$:

$$\Phi(k, \theta_1, \theta_2) = \frac{f\left(\frac{k}{K}; \theta_1, \theta_2\right)}{\sum_{j=0}^K f\left(\frac{j}{K}; \theta_1, \theta_2\right)} \quad (6)$$

Donde $f(\cdot, \theta_1, \theta_2)$ función de densidad de Beta y la normalización garantiza que las ponderaciones suman uno. La densidad Beta en el intervalo (0,1) viene dada por

$$f(z, \theta_1, \theta_2) = \frac{(z)^{\theta_1-1} (1-z)^{\theta_2-1}}{B(\theta_1, \theta_2)} \quad (7)$$

En donde $B(\theta_1, \theta_2)$ es la función beta que cumple con la propiedad de:

$$B(\theta_1, \theta_2) = \frac{\Gamma(\theta_1) \cdot \Gamma(\theta_2)}{\Gamma(\theta_1 + \theta_2)} = \int_0^1 (u)^{\theta_1-1} (1-u)^{\theta_2-1} du \quad (8)$$

Siendo $\Gamma(\cdot)$ la función gamma. Reemplazando (8) en (7),

$$f(z, \theta_1, \theta_2) = \frac{(z)^{\theta_1-1} (1-z)^{\theta_2-1}}{\frac{\Gamma(\theta_1) \cdot \Gamma(\theta_2)}{\Gamma(\theta_1 + \theta_2)}} \quad (9)$$

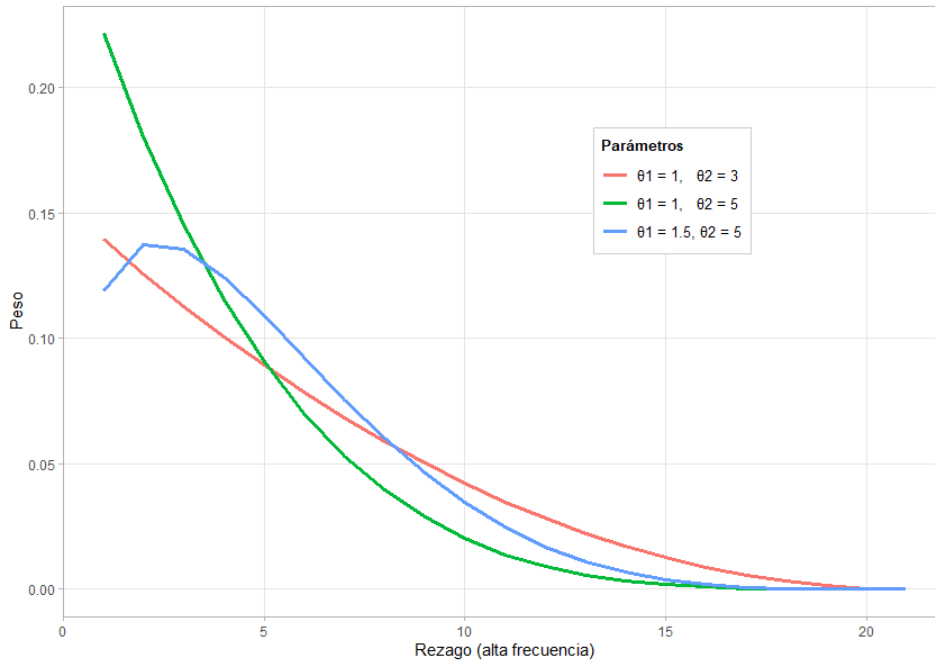
$$f(z, \theta_1, \theta_2) = \frac{(z)^{\theta_1-1} (1-z)^{\theta_2-1} \cdot \Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1) \Gamma(\theta_2)} \quad (10)$$

Con $\theta_1, \theta_2 > 0$ y $z = \frac{k}{K}$. En la figura 1 y 2 se muestran ejemplos de cómo es la ponderación de las observaciones de x_t , dado diferentes funciones y parámetros θ_1 y θ_2 , en el caso de tener datos diarios y una variable dependiente de baja frecuencia mensual. Una ventaja de la función beta son las múltiples formas que puede tomar la función dados los parámetros en $\boldsymbol{\theta}$. En particular, cuando $\theta_1 < \theta_2$ se asigna un peso más alto a las observaciones recientes, si $\theta_1 > \theta_2$ se les asigna un peso bajo a las observaciones anteriores y cuando $\theta_1 = \theta_2 = 1$, se

obtiene exactamente la agregación temporal. En caso de que $\theta_1, \theta_2 > 1$, la función empieza a tener forma de joroba dándole más peso a datos intermedios. De forma alternativa, Ghysels, et al. (2004a) proponen el polinomio exponencial de Almon:

$$\Phi(k, \theta_1, \theta_2) = \frac{e^{\theta_1 k + \theta_2 k^2}}{\sum_{j=0}^K e^{\theta_1 j + \theta_2 j^2}} \quad (11)$$

Figura 1: Funciones de ponderación: Beta normalizada



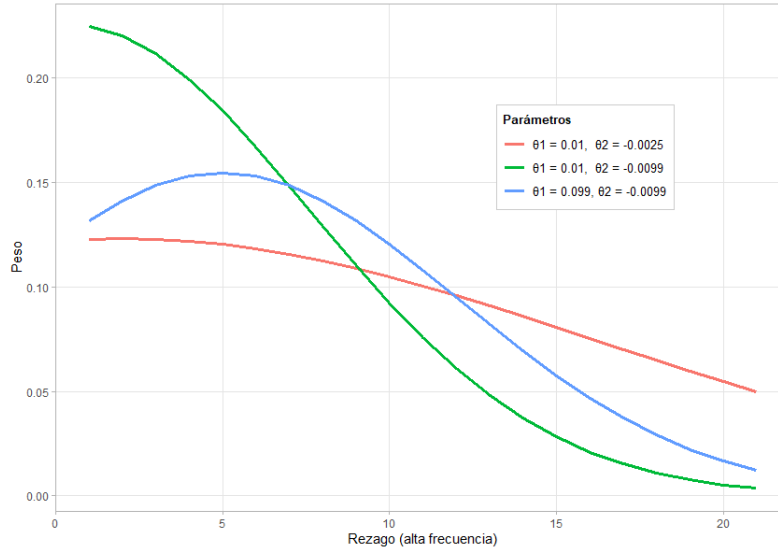
Ghysels et al. (2004a) proponen un orden polinomial de 2 en el exponente. Esta función, al igual que la función beta, asegura que los pesos son positivos, suman 1 y puede generar varias formas la curva de pesos como se muestra en la figura 2. En este caso la forma de la función depende de:

$$k^* = -\frac{\theta_1}{2\theta_2} \quad (12)$$

En aplicaciones MIDAS suele imponerse $\theta_2 < 0$, de modo que la función sea cóncava hacia abajo. En ese caso, si k^* cae dentro del rango de $[1: K]$ la función es creciente hasta k^* para caer después, con $k^* < 1$ la función es decreciente en todo el rango y si $k^* > K$ es creciente en todo el rango, aunque esta última opción es descartable en la práctica. En general, el parámetro θ_1 controla la inclinación de la función y θ_2 su curvatura, de manera que la distribución exacta del peso entre observaciones recientes y antiguas depende conjuntamente

de ambos parámetros a través de k^* . Aunque, en principio, el grado del polinomio puede aumentar capturando patrones complejos, la forma de la función podría presentar comportamientos, por ejemplo, bimodales lo que sería poco habitual o deseado en modelos MIDAS. Ghysels, et al. (2016), proponen otras funciones de agregación en el paquete “midasr” como la Gompertz, Log-Cauchy y Nakagami.

Figura 2: Funciones de ponderación: Almon exponencial normalizada



Como lo comenta Ghysels, et al. (2016) si bien el modelo MIDAS son modelos lineales en cuanto a sus variables, impone restricciones funcionales no lineales lo que da lugar a no linealidades respecto a los parámetros θ . Por lo tanto, la estimación requiere de mínimos cuadrados no lineales (NLS) encontrando los coeficientes que minimizan:

$$(\beta', \theta') = \hat{\gamma} = \underset{\gamma}{\operatorname{argmin}} \sum \left(Y_t - \beta_0 - \beta_1 \sum_{k=0}^K \Phi(k, \theta) \frac{k}{L^m} x_t \right)^2 \quad (13)$$

Esto implica un método iterativo de optimización en el que se evalúa distintas propuestas de θ hasta converger. En consecuencia, dependiendo del orden del polinomio considerado y de la especificación del MIDAS, la regresión arroja los parámetros estimados de las funciones ponderadora y el beta. Ghysels, et al. (2016), apuntan que es necesario existan derivadas de primer y segundo de $\Phi(k, \theta)$ respecto a los hiperparámetros de modo que $(\frac{\partial^2 \Phi(k, \theta)_i}{\partial \theta_i \partial \theta'})$ esté bien definida. Unido a lo anterior, en el paquete “midasr” se evalúan tanto las condiciones de primer y segundo orden asociadas a la estimación NLS, verificando que el gradiente del criterio sea cercano a cero y que el Hessiano sea positivo definido en el punto óptimo. Esto

asegura que el procedimiento de optimización ha llegado a un mínimo local bien determinado de la función objetivo, lo que, bajo los supuestos estándar de la teoría de NLS, permite obtener estimadores consistentes y realizar inferencia estadística sobre los parámetros del modelo.

El estimador beta queda (Ghysels & Qian, 2019):

$$\hat{\beta} = \left(\sum_{t=1}^T x_t(\boldsymbol{\theta})x_t(\boldsymbol{\theta})' \right)^{-1} \left(\sum_{t=1}^T x_t(\boldsymbol{\theta})Y_t \right) \quad (14)$$

Donde $x_t(\boldsymbol{\theta}) = \left[1, \Phi(k, \boldsymbol{\theta})L^{\frac{k}{m}} x_t \right]$ y $\beta = [\beta_0, \beta_1]'$. O,

$$\hat{\beta} = (X(\boldsymbol{\theta})X(\boldsymbol{\theta})')^{-1}X(\boldsymbol{\theta})'Y \quad (15)$$

Siguiendo la lógica de Clements y Galvão (2008) para la elección de valores iniciales de los hiperparámetros en modelos MIDAS, primero se estiman modelos MIDAS univariados para cada regresor de alta frecuencia, y los parámetros estimados de las funciones de ponderación se utilizan como valores iniciales para los θ_i en la estimación conjunta del modelo con múltiples regresores.

De igual manera, se puede considerar la inclusión de componentes autorregresivos en el modelo ya que, según Clements y Galvão (2009), mejora la capacidad predictiva de modelos de variables macroeconómicas:

$$Y_t = \beta_0 + \lambda Y_{t-1} + \beta_1 \sum_{k=0}^K \Phi(k, \boldsymbol{\theta}) L^{\frac{k}{m}} x_t + e_t \quad (16)$$

Donde λ capta la persistencia Y_t . Sin embargo, Ghysels et al. (2007) muestran que esta forma no es apropiada ya que si se reescribe el modelo así:

$$Y_t - \lambda Y_{t-1} = \beta_0 + \beta_1 \sum_{k=0}^K \Phi(k, \boldsymbol{\theta}) L^{\frac{k}{m}} x_t + e_t$$

Es decir,

$$Y_t - \lambda ZY_t = \beta_0 + \beta_1 \sum_{k=0}^K \Phi(k, \boldsymbol{\theta}) L^{\frac{k}{m}} x_t + e_t$$

$$(1 - \lambda Z)Y_t = \beta_0 + \beta_1 \sum_{k=0}^K \Phi(k, \boldsymbol{\theta}) L^{\frac{k}{m}} x_t + e_t$$

Y, por lo tanto,

$$Y_t = \beta_0(1 - \lambda Z)^{-1} + \beta_1 \sum_{k=0}^K (1 - \lambda Z)^{-1} \Phi(k, \boldsymbol{\theta}) L^{\frac{k}{m}} x_t + \tilde{e}_t$$

(17)

Donde $\tilde{e}_t = (1 - \lambda Z)^{-1} e_t$. Generalizando,

$$Y_t = \beta_0 \left(1 - \sum_{i=1}^p \lambda_i Z^i \right)^{-1} + \beta_1 \sum_{k=0}^K \left(1 - \sum_{i=1}^p \lambda_i Z^i \right)^{-1} \Phi(k, \boldsymbol{\theta}) L^{\frac{k}{m}} x_t + \tilde{e}_t$$

(18)

Si se pone atención al polinomio que surge de la interacción de $(1 - \sum_{i=1}^p \lambda_i Z^i)^{-1}$ y $L^{\frac{k}{m}}$ se genera una respuesta estacional inducida de y_t ante x_t aunque este último no tenga realmente un comportamiento estacional. Clements y Galvão (2008, 2009) solucionan este problema al incorporar una restricción de factor común haciendo que la respuesta de Y_t ante x_t se mantenga no estacional:

$$Y_t = \beta_0 + \sum_{i=1}^p \lambda_i Z^i Y_t + \beta_1 \sum_{k=0}^K \left(1 - \sum_{i=1}^p \lambda_i Z^i \right) \Phi(k, \boldsymbol{\theta}) L^{\frac{k}{m}} x_t + e_t$$

(19)

Esta forma funcional es conocida de AR-MIDAS evita la aparición de patrones estacionales espurios y mantiene una estructura de rezagos coherente entre la variable de baja y alta frecuencia. La estimación consiste en estimar una regresión MIDAS sin componente autorregresivo, se toman los residuales estimados \hat{e}_t de, por ejemplo, (4) y se calculan los valores iniciales para λ , digamos $\hat{\lambda}_0$:

$$\hat{\lambda}_0 = (\sum \hat{e}_t^2)^{-1} \sum \hat{e}_t \hat{e}_{t-1}$$

(20)

Luego se construye $Y_t^* = Y_t - \hat{\lambda}_0 Y_{t-1}$ y $x_t^* = x_t - \hat{\lambda}_0 x_{t-1}$ y se vuelve a estimar el MIDAS para obtener un primer estimador de $\boldsymbol{\theta}$. A continuación, se estima un nuevo valor para λ , que podría ser $\hat{\lambda}_1$, de los residuales de la segunda regresión y se repite el procedimiento iterativo de forma que se encuentren los parámetros $\hat{\lambda}$ y $\hat{\boldsymbol{\theta}}$, de forma que se minimicen la sumatoria de cuadrados de residuo.

Ahora, en el caso de tener más variables de diferente frecuencia en el mismo modelo, que, por ejemplo, sea semanales (x_t) y diarias (z_t):

$$Y_t = \beta_0 + \sum_{i=1}^p \lambda_i Z^i Y_t + \beta_1 \sum_{k_1=0}^{K_1} \Phi(k_1, \theta_1) L^{\frac{k_1}{m_1}} x_t + \beta_2 \sum_{k_2=0}^{K_2} \Phi(k_2, \theta_2) L^{\frac{k_2}{m_2}} z_t + e_t \quad (21)$$

Donde K_1 y K_2 son el numero total de rezagos de cada frecuencia.

De forma más general, si se consideran j regresores de alta frecuencia $x_{i,t}^{(m_i)}$, cada uno observado m_i veces dentro del periodo de muestreo de y_t :

$$Y_t = \beta_0 + \sum_{i=1}^p \lambda_i Z^i Y_t + \sum_{i=1}^j \beta_i \sum_{k_i=0}^{K_i} \Phi_i(k_i, \theta_i) L^{k_i/m_i} x_{i,t}^{(m_i)} + e_t \quad (22)$$

Con $K_i = (m_i \cdot n_i) - 1$ y donde, incluso, cada variable explicativa puede tener su propia especificación función de ponderación.

En el caso de U-MIDAS, se elimina la función ponderadora $\Phi(k, \theta)$, por lo que permite que cada rezago tenga un coeficiente asociado (Armesto, et al. 2010). La ecuación equivalente sería:

$$Y_t = \beta_0 + \sum_{k=0}^K \beta_{k+1} L^{\frac{k}{m}} x_t + e_t \quad (23)$$

Como se expuso anteriormente, esta especificación ofrece mayor flexibilidad al no forzar que los rezagos de las variables sigan una forma preespecificada. A diferencia del MIDAS restringido como lo llaman en Foroni, et al. (2012), esta ecuación puede estimarse por MCO, sin embargo, tiene el problema de la proliferación de parámetros a medida que crece el muestreo de alta frecuencia en términos de la variable dependiente (m) y la cantidad de rezagos (n) (Armesto, et al. 2010).

Foroni, et al. (2012) parten de un VAR mensual donde tanto la variable dependiente como la independiente están en la misma frecuencia, pero la primera solo se observa de forma trimestral:

$$a(L)y_t = b_j(L)x_t + e_t$$

(24)

Donde $a(L) = 1 - a_1L - \dots - a_pL^p$ y $b_j(L) = 1 - b_{j,1}L - \dots - b_{j,p}L^p$. Siendo así, hay que definir algún tipo de agregación de y_t para ver Y_t . Sea $w(L)$ ese esquema de agregación temporal que puede ser $w(L) = w_0 + w_1L + w_2L^2 + \dots + w_{m-1}L^{m-1}$ en caso de una variable flujo, $w(L) = 1$ para variables stock o, incluso puede tomar la forma de $w(L) = \frac{1}{m}(w_0 + w_1L + w_2L^2 + \dots + w_{m-1}L^{m-1})$.

El punto de partida de Foroni, et al. (2012) es construir un polinomio de rezagos $\beta(L)$ tal que $\beta(L)a(L) = h(L)$. Premultiplicando $w(L)$ y $\beta(L)$ a (24):

$$a(L)\beta(L)w(L)y_t = b_j(L)\beta(L)w(L)x_t + \beta(L)w(L)e_t$$

(25)

Resolviendo los polinomios y teniendo en cuenta que $h(Z) = h(L^k) = h(L)$ se concluye que, bajo ciertas condiciones, el polinomio $\beta(L)$ usando rezagos de alta frecuencia al interactuar con los otros polinomios $a(L)$, $w(L)$ y $b_j(L)$ hace que (i) las potencias del polinomio $h(L)$ sean múltiplos de m , es decir, que están perfectamente alineadas con la periodicidad de la variable de baja frecuencia; (ii) y_t se agregue como serie de baja frecuencia convirtiéndose en Y_t , (iii) x_t puede expresarse como serie de alta frecuencia e (iv) induce a una autocorrelación serial en los errores de baja frecuencia de manera que el término de error de baja frecuencia puede representarse como un proceso de media móvil (MA). La ecuación resultante es:

$$h(Z)Y_t = \psi(L)x_t + q(Z)u_t$$

(26)

Donde $\psi(L) = b_j(L)\beta(L)w(L)$ es un polinomio de rezagos de alta frecuencia con potencias múltiplos de m , cuyos parámetros dependen, a su vez, de los coeficientes a , w y b de los polinomios $a(L)$, $w(L)$ y $b_j(L)$. Nótese que los parámetros $\psi(L)$ es el que dicta el comportamiento dinámico de x_t , mientras en el MIDAS restringido la dinámica se impone a través de la función con $\Phi(k, \theta)$ de forma que (Foroni, et al. 2012):

$$\psi(L) = \beta_1 \sum_{k=0}^K \Phi(k, \theta) L^{\frac{k}{m}}$$

(27)

Por tanto, el MIDAS restringido se obtiene como caso particular al imponer que los coeficientes sigan la curvatura de una función ponderadora $\Phi(k, \theta)$. En este sentido, el modelo MIDAS restringido queda anidado dentro del U-MIDAS (Foroni, et al. 2012).

A continuación, se presenta un ejemplo del modelo U-MIDAS en forma matricial. Para ello, de igual manera, se debe considerar una serie de baja frecuencia como puede ser una variable trimestral Y_t de tamaño T y una serie de alta frecuencia mensual x_t de tamaño $3T$. Suponiendo $m = 3$, para construir la matriz X :

1. Decidir n de manera de que se obtenga K . En este ejemplo, $n = 2$ y, por tanto, $K = 2 \cdot 3 - 1 = 5$
2. Eliminar los primeros $n - 1$ trimestres de Y_t para que se tengan disponibilidad de rezagos.
3. Construir filas de x_t usando $K + 1$ valores hacia atrás desde $3T$

Teniendo en cuenta la tabla 1:

Tabla 1: Alineación de datos mensuales y mensuales

Trimestre (T)	Meses ($3T$)
1	1, 2, 3
2	4, 5, 6
3	7, 8, 9
4	10, 11, 12
5	13, 14, 15

La forma matricial (ignorando el intercepto) quedaría:

$$\begin{bmatrix} Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{bmatrix} = \begin{bmatrix} x_6 & x_5 & x_4 & x_3 & x_2 & x_1 \\ x_9 & x_8 & x_7 & x_6 & x_5 & x_4 \\ x_{12} & x_{11} & x_{10} & x_9 & x_8 & x_7 \\ x_{15} & x_{14} & x_{13} & x_{12} & x_{11} & x_{10} \end{bmatrix} \text{ donde } \begin{matrix} 3T = 6 \\ 3T = 9 \\ 3T = 12 \\ 3T = 15 \end{matrix} \quad (28)$$

En consecuencia, el modelo quedaría como

$$\begin{bmatrix} Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \end{bmatrix} = \begin{bmatrix} x_6 & x_5 & x_4 & x_3 & x_2 & x_1 \\ x_9 & x_8 & x_7 & x_6 & x_5 & x_4 \\ x_{12} & x_{11} & x_{10} & x_9 & x_8 & x_7 \\ x_{15} & x_{14} & x_{13} & x_{12} & x_{11} & x_{10} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \end{bmatrix} + \begin{bmatrix} e_2 \\ e_3 \\ e_4 \\ e_5 \end{bmatrix} \quad (29)$$

Sin pérdida de generalidad, los modelos MIDAS pueden incorporar factores derivados de modelos de factores dinámicos (DFM). Dentro de las ventajas que tiene esta metodología es la robustez frente a inestabilidades de variables individuales (Andreou, et al. 2010b), la reducción de dimensionalidad y la mitigación de problemas de multicolinealidad al evitar incluir simultáneamente muchos indicadores que contienen información muy similar y, por

lo tanto, evitando sesgos en los estimadores. Contrario a eso los DFM, encuentran relaciones no observadas entre las variables y las condensada en pocas variables resultantes. Esto lleva a reducir el costo computacional de la estimación, tener coeficientes insesgados, más eficientes y, por lo tanto, significativos, aunque con pérdida en la interpretación económica de los factores estimados. Siguiendo Marcellino y Schumacher (2010), el modelo de factores puede escribirse como:

$$\mathbf{X}_t = \mathbf{\Lambda}F_t + \boldsymbol{\varepsilon}_t \quad (30)$$

$$F_t = \sum_{j=1}^p \boldsymbol{\Omega}_j F_{t-j} + \boldsymbol{\eta}_t \quad (31)$$

La ecuación (30) es la ecuación de medición donde \mathbf{X}_t es un vector de variables observadas en t , $\mathbf{\Lambda}$ es una matriz de loadings, F_t es un vector de factores observados en t , mientras la ecuación (31) es la de transición que modela la dinámica de los factores como un VAR de orden p . En este caso, la frecuencia de \mathbf{X}_t y F_t es la misma. A diferencia de Andreou, et al. (2010b), los loadings ($\mathbf{\Lambda}$) son estáticos en el tiempo lo que simplifica la estimación de los factores. Una vez, estimados los factores se emplean en la ecuación MIDAS restringido convirtiéndose en Factor-MIDAS:

$$Y_t = \beta_0 + \sum_{i=1}^p \lambda_i Z^i Y_t + \sum_{i=1}^j \beta_i \sum_{k_i=0}^{K_i} \Phi_i(k_i, \boldsymbol{\theta}_i) L^{k_i/m_i} \hat{F}_{i,t}^{(m_i)} + e_t \quad (32)$$

O en la de U-MIDAS:

$$h(Z)Y_t = \psi(L)\hat{F}_t + q(Z)u_t \quad (33)$$

Nótese que, al igual que Andreou, et al. (2010b), es posible tener un componente autorregresivo de la variable de pendiente sin necesidad de que sea factor. En este caso, se consideró el componente autorregresivo en las mismas unidades de la variable dependiente pero las demás explicativas de la misma frecuencia si fueron sometidas a factores dinámicos, así como las series semanales y diarias, que se describen en la siguiente sección.

4. Datos y Preprocesamiento

El objetivo de esta sección es presentar los datos usados, así como las transformaciones aplicadas para la estimación de los modelos MIDAS.

Para la predicción de inflación se tuvieron en cuenta 30 variables en total, incluyendo el índice de precios al consumidor (IPC). Se tienen 15 variables mensuales, 3 semanales y 13 diarias. La selección preliminar de las variables se hace a la luz de la teoría económica procurando abarcar las diferentes fuentes de presiones o frenos inflacionarios ya sea vía oferta, demanda, monetaria, laboral u externas. Adicionalmente, se tomaron en cuenta variables de artículos cuyo foco es justamente pronóstico de inflación para Colombia como Cardenas, et al. (2020) o Martínez, et al. (2023). La base de datos original presenta variables en niveles desde enero de 2006 hasta septiembre de 2025. En total se tienen 237 meses, 1030 semanas y 4800 días. La tabla 2 muestra las variables seleccionadas, su frecuencia y su fuente.

Tabla 2: Variables

Variable	Frecuencia	Fuente
Índice de precios al consumidor (IPC) (Colombia)	Mensual	DANE
Ocupados	Mensual	DANE
Índice de precios al productor (IPP)	Mensual	DANE
Índice de precios de vivienda nueva (IPVNR)	Mensual	Banco de la República
Reservas Internacionales	Mensual	Banco de la República
Índice de tasa de cambio real (ITCR)	Mensual	Banco de la República
Índice de términos de intercambio	Mensual	Banco de la República
Índice de producción industrial (IPI)	Mensual	Banco de la República
Balanza Comercial	Mensual	Banco de la República
Tasa de política monetaria	Mensual	Banco de la República
Índice de precios al consumidor (Estados Unidos)	Mensual	FRED St. Louis
Déficit Fiscal	Mensual	Ministerio de Hacienda
Índice Oceánico del Niño (ONI)	Mensual	NAOO
Índice de precios de los alimentos	Mensual	FAO
Prestamos no rentables	Mensual	Bloomberg
M2	Semanal	Banco de la República
Tasa de crédito de consumo	Semanal	Banco de la República
Tasa de crédito preferencial	Semanal	Banco de la República

Colcap	Diaria	Bloomberg
Tasa de cambio EUR/COP	Diaria	Bloomberg
Tasa de cambio USA/COP	Diaria	Bloomberg
Índice S&P GSCI trigo	Diaria	Bloomberg
Índice S&P GSCI soya	Diaria	Bloomberg
Índice S&P GSCI algodón	Diaria	Bloomberg
Índice S&P GSCI café	Diaria	Bloomberg
Índice S&P GSCI ganado bovino	Diaria	Bloomberg
Índice S&P GSCI oro	Diaria	Bloomberg
Expectativas de inflación a 1 año	Diaria	Cálculos propios a partir de la información del Banco de la República
VIX	Diaria	Bloomberg
S&P 500	Diaria	Bloomberg
Precio del petróleo Brent	Diaria	EIA
Bovespa	Diaria	Bloomberg

En cuanto a las variables mensuales, vale la pena resaltar las reservas internacionales que inciden en la cantidad de dinero circulante y, según Mirza, et al. (2024), es un indicador olvidado a la hora de hacer pronósticos de inflación pese a su elevado poder predictivo. La balanza comercial se obtiene de la diferencia entre las exportaciones y las importaciones. Por su parte, se incluye el déficit fiscal en línea con la teoría fiscal del nivel de precios (Bianchi, et al. 2023) en la que el balance fiscal puede tener efectos inflacionarios, pues, como dice Sargent (2013), una alta inflación persistente es siempre y en todo lugar un fenómeno fiscal. Por último, el índice ONI que identifica eventos cálidos (El Niño) y fríos (La Niña) en el océano Pacífico tropical y puede tener impacto en el precio de los alimentos a través de choques de oferta (Cardenas, et al. 2020).

Las series semanales son el agregado monetario M2 que comprende el efectivo en circulación, depósitos bancarios a la vista y los depósitos a plazo; las tasas de interés de crédito de consumo y preferencial, determinantes del consumo de los hogares y la inversión de las empresas. En el caso de las variables diarias se toma como universo los días bursátiles para Colombia, se incluyen las tasas de cambio de las divisas más importantes para Colombia y el mundo como el dólar estadounidense y el euro por su efecto “pass-through” sobre los precios internos. Las expectativas de inflación se aproximan mediante la tasa de los bonos

cupón cero tasa fija y variable a 1 año. Los índices precios de commodities de Standard & Poors (S&P) como el índice S&P GSCI trigo son considerados dado que Chen, et al. (2014) muestran que en países exportadores y con esquemas inflación objetivo estas variables tienen poder predictivo en la inflación. Asimismo, ninguna serie fue deflactada o incluida en términos reales en caso de que existan las versiones nominal y real; y se tomaron las series desestacionalizadas.

Un aspecto importante antes de la estimación de los modelos es la alineación de los datos. Como se mencionó, la cantidad de meses son 237, mientras que las semanas ascienden a 1030, es decir, cerca de 4,35 semanas por mes. Sin embargo, en el paquete “midasr” m debe ser un entero, por lo que se deben considerar 4 o 5 semanas. Se optó por meses de 4 semanas y, cuando se presenta el caso de meses con 5 semanas se toma el promedio entre la cuarta y la quinta semana. De forma análoga, se hizo para las variables diarias. Dado el número de festivos que hay en Colombia comparado con el resto del mundo, los días bursátiles por mes es, en promedio 20,25. Se definió entonces 21 días por mes. En los meses con 21 días bursátiles se utilizan las 21 observaciones; en aquellos con solo 20 días, el día 21 se asigna igual al valor del día 20, de modo que todos los meses queden representados por 21 observaciones diarias. Tras estos ajustes, las observaciones semanales quedaron en 944 y los diarios pasaron a 4956.

Posteriormente, se procede a volver las series estacionarias aspecto importante en estos modelos según Marcellino y Schumacher (2010), Andreou, et al. (2010b) y Ghysels, et al. (2004a). Para esto, teniendo en cuenta que hay variables con datos negativos como el déficit fiscal se aplica logaritmo con el ajuste del seno hiperbólico que permite tratar de forma coherente valores positivos y negativos. Luego se sacan las primeras diferencias para eliminar tendencias. El procedimiento fue el mismo para todas las series de las diferentes frecuencias. Así, se debería tener la inflación mensual en el caso de la dependiente y las variables explicativas estacionarias.

5. Estimación y Resultados

Esta sección muestra procedimientos, criterios, estimación y análisis de los resultados obtenidos de los modelos de factores dinámicos y MIDAS.

Inicialmente, antes de estimar el MIDAS, se procede a obtener los factores del modelo de factores dinámicos. Para esto, se calculan diferentes modelos para cada tipo de datos, es decir, se estiman unos factores para las variables mensuales, otros diferentes para las semanales y lo mismo para las diarias. Se toman las series con la transformación logarítmica y en primeras diferencias para que los factores sean de igual manera estacionarios. Los DFM requieren que las series estén estandarizadas por lo que, se les aplica esta transformación. Posteriormente, se realiza la elección la cantidad de componentes principales a tener en cuenta de forma visual y la cantidad de rezagos tomando en cuenta los criterios de información de forma que el modelo DFM presente convergencia.

En el caso de los datos mensuales, consideran 6 componentes principales (figura 3¹) y 3 rezagos sugeridos por los criterios de información AIC y FPE. La figura 4 muestra las correlaciones de los componentes principales con las variables con las que se construyeron. El primer componente principal parece estar recogiendo el comportamiento de precios externos ya que tiene alta correlación con los términos de intercambio, el índice de precios de los alimentos de la FAO y la inflación de Estados Unidos. El segundo está muy correlacionado con la tasa de política monetaria y el IPP. El tercer y quinto factor parecen estar relacionados con la política fiscal y monetaria, el cuarto con la economía real y el comercio internacional, y el sexto casi que exclusivamente tiene que ver con el fenómeno del niño o la niña.

Para los datos semanales, se reducen la dimensionalidad de 3 variables a 2 factores (figura 5) con 10 rezagos según los criterios AIC, HQ y FPE. Al mirar las correlaciones (figura 6), el primer factor recoge las presiones inflacionarias por el lado monetario, mientras que el segundo esta más relacionado con fenómenos que alivian la inflación vía tasas de interés de consumo. Los factores dinámicos no contemplan información de la tasa de interés preferencial.

En el caso de las variables diarias, se eligieron 4 componentes principales (figura 7) con 6 rezagos según los criterios AIC y FPE. El primer factor carga los mercados financieros, así como el riesgo, el segundo commodities agrícolas, el tercero las monedas referencia y el cuarto está relacionada exclusivamente con el precio del petróleo WTI (figura 8).

Una vez estimado los factores dinámicos se procede a la estimación del MIDAS. Con el fin de ahorrar tiempo y recursos computacionales en la estimación y tener un punto de referencia en la cantidad de rezagos se estiman modelos ARIMAX con dos rezagos de la inflación y un solo factor mensual al tiempo y considerando hasta 12 rezagos. La elección de la cantidad de rezagos de cada factor se hizo teniendo en cuenta el criterio AIC². En concreto, se tomó en cuenta el primer y segundo rezago del factor 1, el primero del factor 2, del segundo al cuarto del factor 3, el primero del factor 4, del primero hasta el tercero del factor 5 y únicamente el primero del factor 6.

Para los factores semanales, se aplica un procedimiento similar pero no para determinar la cantidad de rezagos sino la especificación y la función ponderadora apropiada, así como la inserción o no de la restricción del factor común. En ambos factores, la especificación que mejor criterio AIC tuvo fue la que considera la función ponderadora del polinomio de Almon, el caso de agregación A de Ghysels, et al. (2016) y sin el ajuste de Clements y Galvão (2008).

Teniendo lo anterior en cuenta y siguiendo a Andreou, et al. (2010b), se estiman diferentes modelos MIDAS con factores mensuales, semanales e incorporando un solo factor dinámico

¹ Figuras 3 en adelante en el anexo

² Hiperparámetros y funciones en tabla 5 del anexo

diario al tiempo principalmente para evitar la proliferación de parámetros en los modelos U-MIDAS de forma que se pueda hacer el test hAh, que permite decidir si la restricción de la función ponderadora es adecuada o no por lo que se debe considerar un U-MIDAS. La grilla considerada tomó en cuenta hasta 3 meses de historia en cada frecuencia, la inclusión o no de la restricción de factor común y 3 tipos de agregación.

Posteriormente, se depuraron los modelos que no alcanzaron convergencia y aquellos que no alcanzaron un mínimo en su función. Los modelos que quedaron se dividieron en dos grupos: aquellos donde el test hAh validaba la restricción de parámetros y los que no. Quedaron 12 modelos MIDAS y 18 U-MIDAS entre los que se eligió el mejor de cada grupo de acuerdo con el criterio AIC y se evaluó sus RMSE in-sample y out-of-sample tomando como modelo de referencia un AR(1). La muestra se dividió 70% in-sample y 30% out-of-sample.

La parte superior de la tabla 3, muestra el RMSE de cada modelo in-sample y out-of-sample mientras que la parte de abajo muestra esa misma métrica de error relativa al AR(1) y entre los modelos MIDAS. In-sample, los resultados muestran que tanto el modelo MIDAS como el U-MIDAS presentan mejoras en la predicción frente al AR(1) en un 10% y un 44%, siendo así, el U-MIDAS mejor modelo en esta muestra resultado similar a Foroni, et al. (2012). Sin embargo, out-of-sample el modelo ganador es el MIDAS comparado con el U-MIDAS, lo que sugiere un sobreajuste del U-MIDAS; y el AR(1), aunque la ganancia predictiva frente al modelo de referencia se reduce a la mitad lo que es consistente con lo encontrado por Armesto, et al. (2010) y Andreou, et al. (2010a) que remarcan que el AR(1) es un benchmark bastante fuerte haciendo que en algunos casos no exista diferencia significativa.

Tabla 3: Resultados mejor MIDAS, mejor U-MIDAS y AR(1)

Modelo	RMSE	
	In-sample	Out-of-sample
Mejor MIDAS	0,002	0,003
Mejor U-MIDAS	0,001	0,004
AR(1)	0,002	0,003
MIDAS/AR(1)	0,904	0,951
U-MIDAS/AR(1)	0,569	1,330
U-MIDAS/MIDAS	0,630	1,398

Con respecto a la combinación de pronósticos, toma la siguiente forma:

$$\hat{Y}_t^{comb} = \sum_{q=1}^Q w_q \hat{Y}_t^q$$

(34)

Donde \hat{Y}_t^q es la predicción de \hat{Y}_t del modelo q de forma individual, w_q es la ponderación que se le da al modelo q con $\sum w_q = 1$; \hat{Y}_t^{comb} es la predicción conjunta de los diferentes modelos ponderados y Q es la cantidad total de modelos tomados en cuenta. Torres (2015) comenta que, aunque es clara la evidencia sobre los beneficios de combinar diferentes modelos, la discusión está en la forma de ponderar estos pronósticos. En este caso se consideran 4 formas: promedio simple, pesos basados en el criterio BIC, pesos según MSFE y el Discounted MSFE. Así pues, por ejemplo, cada una de las predicciones de los 12 modelos MIDAS tendrán una ponderación dentro de \hat{Y}_t^{comb} , dependiendo de las métricas que se tomen en consideración.

Donde el promedio simple es:

$$w_q = \frac{1}{Q}$$

(35)

Los pesos basados en el criterio BIC que ponderan de mejor manera los modelos más parsimoniosos:

$$w_q = \frac{\exp\left(-\frac{1}{2} \cdot BIC_q\right)}{\sum_{r=1}^Q \exp\left(-\frac{1}{2} \cdot BIC_r\right)}$$

(36)

Los pesos según MSFE que les asigna mayor carga a aquellos modelos con menor error de predicción:

$$MSFE_q = \frac{1}{T} \sum_t e_{q,t}^2$$

(37)

$$w_q = \frac{1}{MSFE_q}$$

(38)

Y el discounted MSFE, donde los errores más recientes tienen mayor peso que los antiguos:

$$DMSFE_q = \sum_t \lambda^{T-t} \cdot e_{q,t}^2$$

(39)

Con $0 < \lambda < 1$. La tabla 4 compara el RMSE de la combinación de pronósticos de los modelos previamente seleccionados de MIDAS y U-MIDAS, con los pronósticos individuales anteriormente presentados y el AR(1). La forma de ponderación que menor RMSE out-of-sample tuvo fue los pesos basados en el criterio BIC. En general, se tiene que la combinación de pronósticos en el caso del MIDAS no hay ganancias adicionales, mientras que en el caso del U-MIDAS hay una leve mejora. Asimismo, ambas combinaciones de pronósticos le ganan al AR(1) y a pronósticos individuales en especial en el caso del MIDAS, similar a lo encontrado por Torres (2015) y Andreou, et al. (2010b).

Tabla 4: Resultados combinación pronósticos de MIDAS, U-MIDAS y AR(1)

Modelo	RMSE
	Out-of-sample
Mejor MIDAS	0,003
Mejor U-MIDAS	0,004
Combinación MIDAS	0,003
Combinación U-MIDAS	0,003
AR(1)	0,003
MIDAS/AR(1)	0,951
U-MIDAS/AR(1)	1,330
U-MIDAS/MIDAS	1,398
Combinación MIDAS/ Mejor MIDAS	1,010
Combinación U-MIDAS/ Mejor U-MIDAS	0,958
Combinación MIDAS/ AR(1)	0,961
Combinación U-MIDAS/ AR(1)	0,920

Los resultados obtenidos sugieren varias conclusiones interesantes. Como lo muestra Andreou, et al. (2010a), cuando el proceso es persistente, la dinámica del proceso es muy compleja el MIDAS tiende a tener mejores resultados que el AR(1). Armesto, et al. (2010) y Clements y Galvão (2008) confirman lo anterior diciendo que no hay ganancias sistemáticas del MIDAS sobre el AR(1). En este caso, la inflación mensual parece no ser un proceso muy persistente ni del todo complejo razón por la cual también el U-MIDAS no tuvo un gran desempeño out-of-sample (Foroni, et al. 2012) mientras que el MIDAS sí. Respecto a la combinación pronósticos, aunque al comparar las métricas de error no hay una ganancia muy significativa frente a los pronósticos individuales, al mirar la figura 9 y 10 que muestran que la combinación de pronósticos se ve que, en especial para el U-MIDAS, se obtienen predicciones más estables (menos volátiles).

6. Conclusiones

La diferencia de muestreo en las series económicas y financieras siempre ha sido un problema para los pronosticadores que quieren incluir estas series en un mismo modelo. La práctica tradicional ha sido agregar o promediar las series para que queden de baja frecuencia. Una alternativa a esto son los modelos tipo MIDAS que propone una forma parsimoniosa, consistente y eficiente de unir en un mismo modelo datos de diferente frecuencia imponiendo una función ponderadora de pocos parámetros. Este trabajo se estimaron diferentes regresiones MIDAS para modelar la inflación mensual en Colombia. Los resultados muestran que (i) la inflación parece ser un proceso poco persistente y que puede modelarse con especificaciones parsimoniosas, (ii) el MIDAS restringido presenta leves mejoras en la predicción si se considera un pronóstico individual, (iii) la combinación de pronósticos tiene beneficios en el caso del U-MIDAS frente a pronósticos individuales y, (iv) en el caso de que se compare contra el AR(1), se obtienen predicciones más precisas y menos volátiles, aunque el modelo autorregresivo es un competidor importante.

Dentro de las posibles mejoras está la exploración de más variables explicativas en especial la incorporación de un mecanismo que recoja el esquema de inflación objetivo del banco central, una mejor modelación de quiebres estructurales como la pandemia y el paro nacional de 2021 que quedan en periodo de out-of-sample, considerar loadings dinámicos en el modelo de factores dinámicos como en Andreou, et al. (2010b), explorar otras especificaciones a la hora de sintonizar los parámetros y extender a un MIDAS con adelantos.

7. Referencias

- Andreou, E., Ghysels, E., & Kourtellos, A. (2010a). Regression models with mixed sampling frequencies. *Journal of Econometrics*, 158(2), 246-261.
- Andreou, E., Ghysels, E., & Kourtellos, A. (2010b). Should macroeconomic forecasters use daily financial data and how?. *Journal of Business & Economic Statistics*, 31(2), 240-251.
- Armesto, M. T., Engemann, K. M., & Owyang, M. T. (2010). Forecasting with mixed frequencies. *Federal Reserve Bank of St. Louis Review*, 92(6), 521-536.
- Bai, J., Ghysels, E., & Wright, J. H. (2010). State space models and MIDAS regressions. *Econometric Reviews*, 32(7), 779-813.
- Bianchi, F., Faccini, R., & Melosi, L. (2023). A fiscal theory of persistent inflation. *The Quarterly Journal of Economics*, 138(4), 2127-2179.
- Chen, Y. C., Turnovsky, S. J., & Zivot, E. (2014). Forecasting inflation using commodity price aggregates. *Journal of Econometrics*, 183(1), 117-134.
- Clements, M. P., & Galvão, A. B. (2008). Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States. *Journal of Business & Economic Statistics*, 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.
- Cárdenas-Cárdenas, J. A., Caicedo-García, E., González-Molano, E. R., Edgar, C. G., & Eliana, R. (2020). Estimación de la variación del precio de los alimentos con modelos de frecuencias mixtas. *Borradores de Economía*; No. 1109.
- Foroni, C., Marcellino, M., & Schumacher, C. (2012). Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 178(1), 57-82.
- French, K. R., Schwert, G. W., & Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of financial Economics*, 19(1), 3-29.
- Galeano-Ramírez, F. J., Martínez-Cortés, N., & Rojas-Martínez, C. D. (2021). Nowcasting colombian economic activity: Dfm and factor-midas approaches. *Borradores de Economía*; No. 1168.
- Ghysels, E., & Qian, H. (2019). Estimating MIDAS regressions via OLS with polynomial parameter profiling. *Econometrics and statistics*, 9, 1-16.
- Ghysels, E., Kvedaras, V., & Zemlys, V. (2016). Mixed frequency data sampling regression models: The R package midasr. *Journal of statistical software*, 72, 1-35.
- Ghysels, E., Santa-Clara, P., & Valkanov, R. (2004a). There is a risk-return trade-off after all. *Journal of Financial Economics*, 76(3), 509-548.
- Ghysels, E., Santa-Clara, P., & Valkanov, R. (2004b). The MIDAS touch: Mixed data sampling regression models.
- Ghysels, E., Sinko, A., & Valkanov, R. (2007). MIDAS regressions: Further results and new directions. *Econometric reviews*, 26(1), 53-90.
- González-Molano, E. R. (2010). Bayesian model averaging: an application to forecast inflation in Colombia.
- Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A contribution to the empirics of economic growth. *The quarterly journal of economics*, 107(2), 407-437.
- Marcellino, M., & Schumacher, C. (2010). Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP. *Oxford Bulletin of Economics and Statistics*, 72(4), 518-550.
- Martínez-Rivera, W., González-Molano, E. R., & Caicedo-García, E. (2023). Forecasting Inflation from Disaggregated Data: The Colombian case. *Borradores de Economía*; No. 1251.
- Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica: Journal of the Econometric Society*, 867-887.
- Mirza, N., Rizvi, S. K. A., Naqvi, B., & Umar, M. (2024). Inflation prediction in emerging economies: Machine learning and FX reserves integration for enhanced forecasting. *International Review of Financial Analysis*, 94, 103238.

- Sargent, T. J. (2013). Letter to Another Brazilian Finance Minister, 1993. Rational Expectations and Inflation.
- Torres, D. J. (2015). Evaluando la capacidad predictiva del MIDAS para la Eurozona, Italia, Alemania, Francia y Portugal.

8. Anexos

Figura 3: Varianza explicada por componentes principales. Datos mensuales

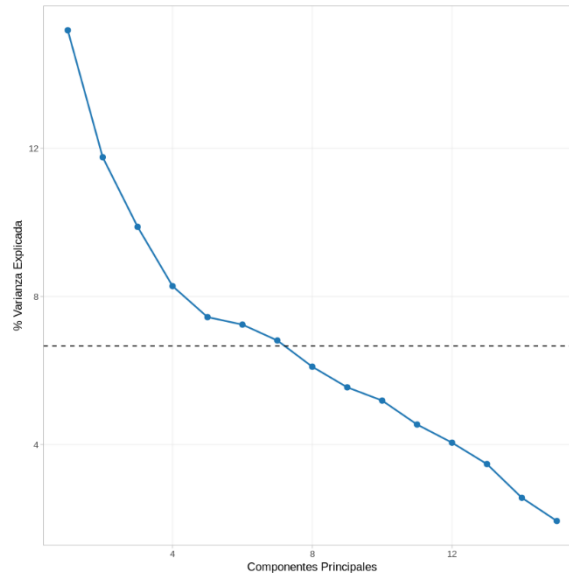


Figura 4: Matriz de correlaciones factores dinámicos y variables. Datos mensuales

Correlaciones: Factores (filas) vs resto de variables (columnas)

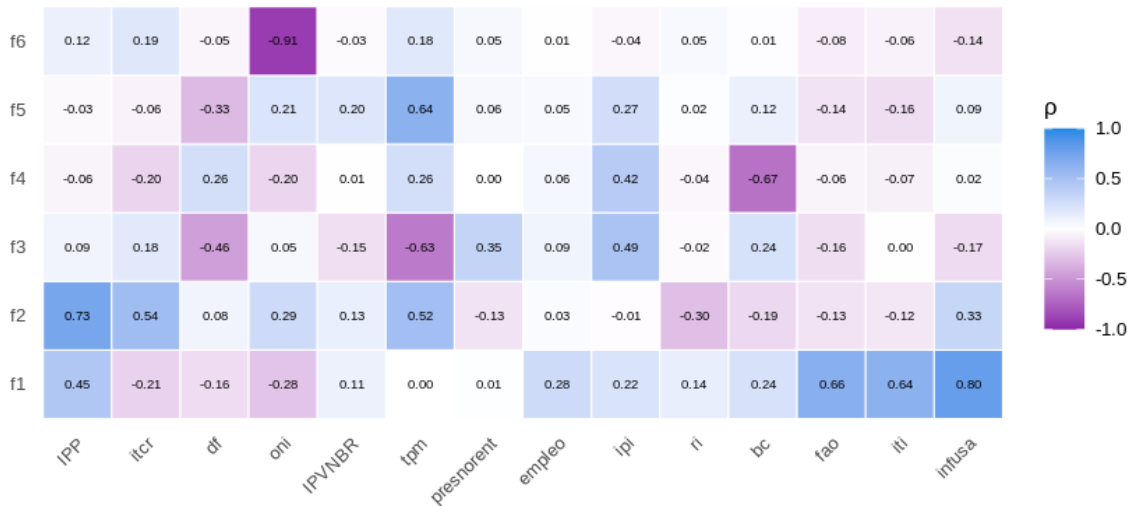


Figura 5: Varianza explicada por componentes principales. Datos semanales

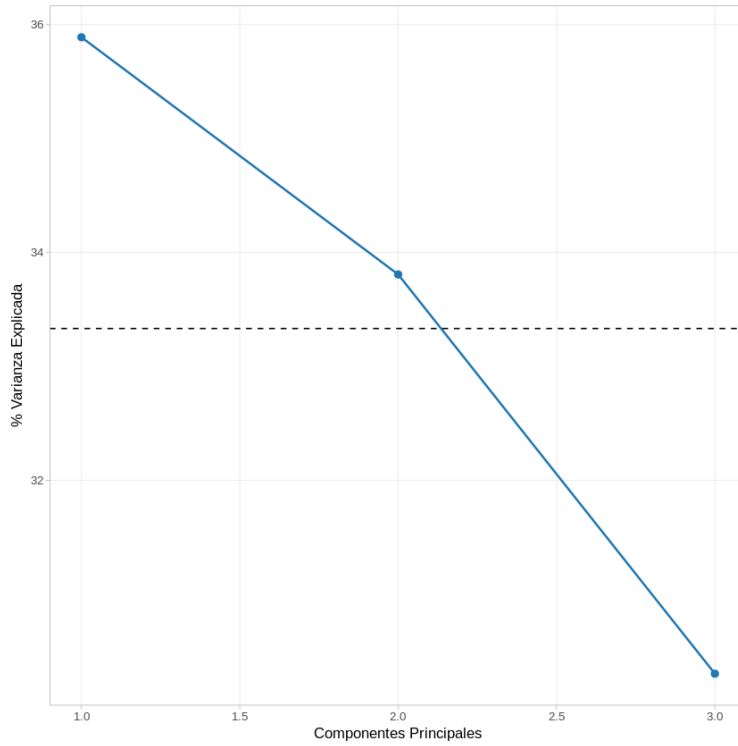


Figura 6: Matriz de correlaciones factores dinámicos y variables. Datos mensuales

Correlaciones: Factores (filas) vs resto de variables (columnas)

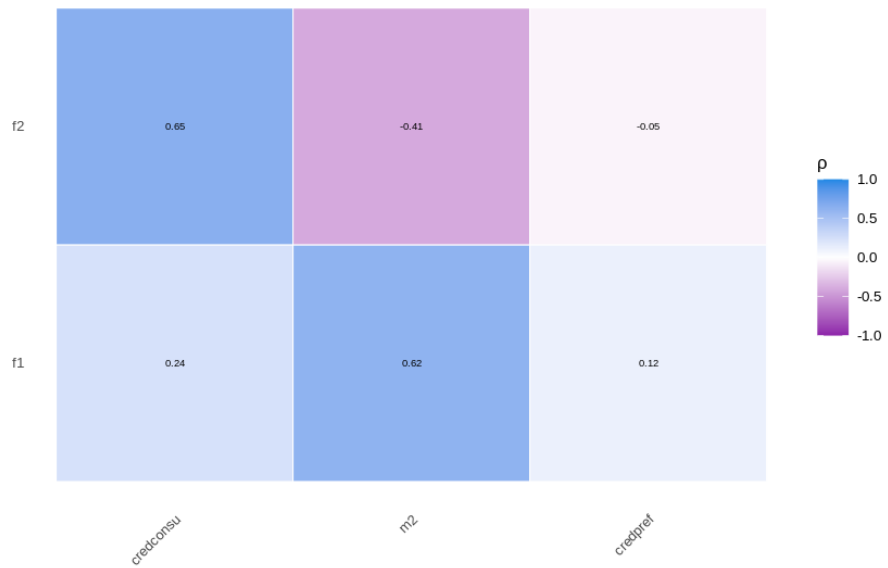


Figura 7: Varianza explicada por componentes principales. Datos diarios

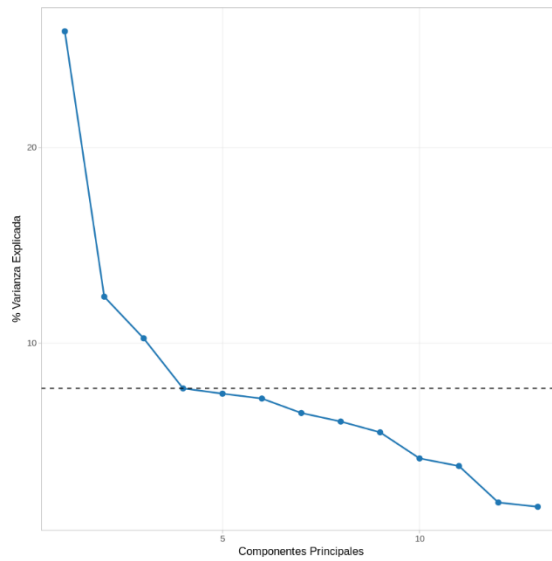


Figura 8: Matriz de correlaciones factores dinámicos y variables. Datos mensuales

Correlaciones: Factores diarios (filas) vs variables (columnas)

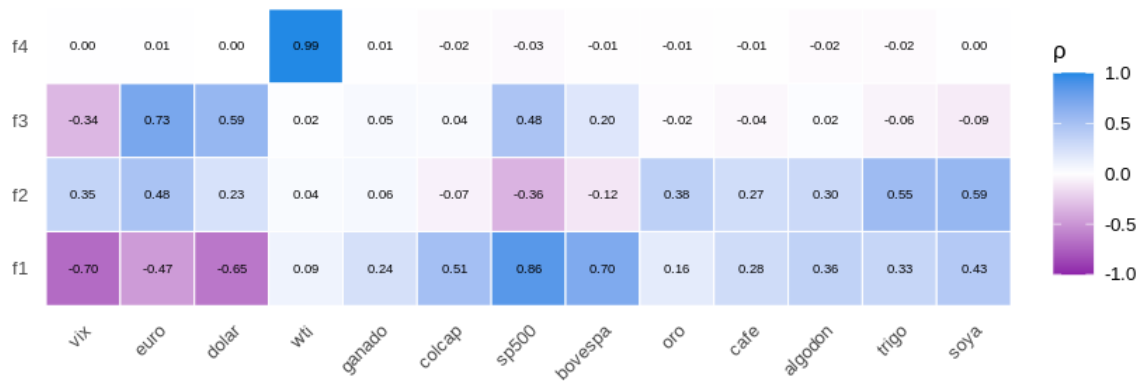


Figura 9: Predicción de inflación out-of-sample. Mejor MIDAS, mejor U-MIDAS y AR(1)

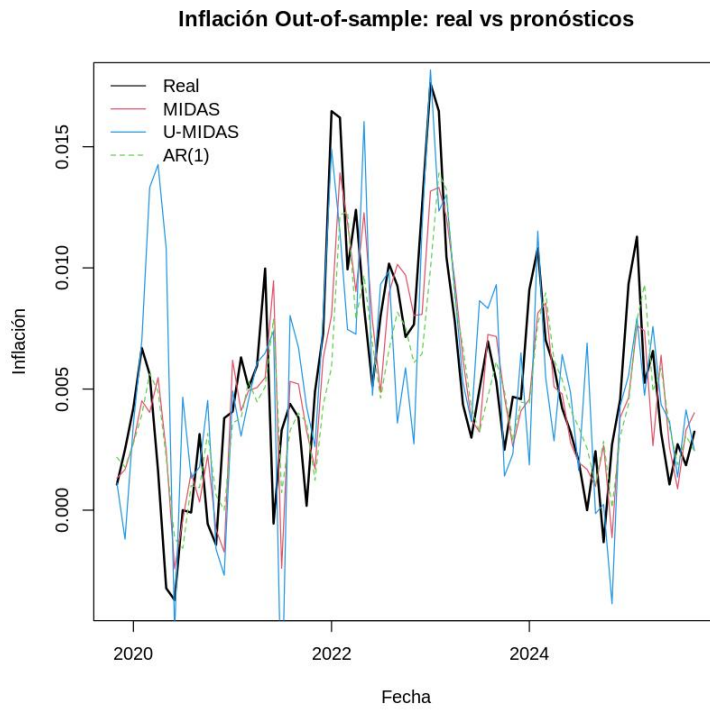


Figura 10: Predicción de inflación out-of-sample. Combinación de pronósticos

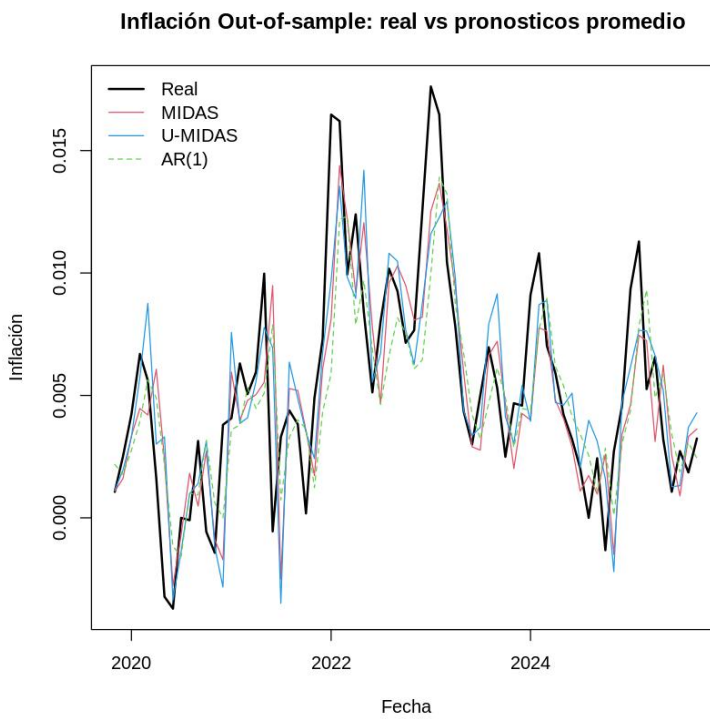


Tabla 5: Grilla de hiperparámetros MIDAS

Set	Hiperparámetro	Valor /Valor inicial
Mensual	Restricción de factor común	“*”/ “ ”
	From	1
	To	12
Semanal	Restricción de factor común (factor mensual)	“*”/ “ ”
	Funciones ponderadoras	Nealmon,nbeta, gompertz
	Tipo	A, B, C
	From	0
	To	(1,3)
	m	4
	Nealmon	0.099, -0.0099
	Nbeta	1, 5, 4
	Gompertzp	0.1, 0.1, 0.1
Diario	Restricción de factor común (factor mensual)	“*”/ “ ”
	Funciones ponderadoras	Nealmon,nbeta, gompertz
	Tipo	A, B, C
	From	0
	To	(1,3)
	m	21
	Nealmon	0.1, -0.0099
	Nbeta	1, 5, 21
	Gompertzp	0.1, 0.1, 0.1