

Proyecciones lineales de alta dimensionalidad

Parte I

Fernando Arias-Rodríguez

Banco Central de Bolivia

29 de agosto de 2024



- ① Introducción
- ② Ecuaciones *Bridge*
- ③ Regresiones MIDAS
- ④ Regresiones UMIDAS
- ⑤ Extensiones Modelos MIDAS

- ① Introducción
- ② Ecuaciones *Bridge*
- ③ Regresiones MIDAS
- ④ Regresiones UMIDAS
- ⑤ Extensiones Modelos MIDAS

- Cuando se tiene información en diferentes frecuencias, la aproximación más sencilla es agregar la información de alta frecuencia para obtener una base de datos balanceada en la baja frecuencia.
- Por ejemplo, se puede tener en la misma base de datos el Producto Interno Bruto (PIB) en frecuencia trimestral, un Índice de Producción Industrial y una Tasa de Desempleo en frecuencia mensual.
- Lo usual es convertir las series mensuales en trimestrales, para tener todas las variables en frecuencia trimestral.

- Sin embargo, la agregación temporal genera una pérdida de información.
- Incluso, podría alterar las propiedades estocásticas de la serie.
- Ello conllevaría a estimar modelos espurios con series agregadas o que predicen de manera diferente cuando se agrega y cuando se trabaja en la frecuencia original.
- La alternativa es trabajar con modelos que sean capaces de mezclar información en diferentes frecuencias.

- La existencia de información en frecuencias mixtas se relaciona directamente con el concepto de *Nowcasting*.
- Se consideran algunas alternativas para modelar información en frecuencias mixtas:
 - Ecuaciones *Bridge*.
 - *Mixed-data Sampling Models* (MIDAS).

Se deben tener en cuenta algunas consideraciones:

- La elección de indicadores es relevante en el contexto de *Nowcasting*, es decir, diferentes indicadores pueden ser usados en tramos diferentes del tiempo (según la información disponible, la fase del ciclo, entre otras).
- Revisiones y tratamientos previos de la información (desestacionalización, remoción de datos atípicos) juegan un papel importante en el resultado final.
- La evaluación de los modelos debe hacerse con información en tiempo real.

1 Introducción

2 Ecuaciones *Bridge*

Notación

Caso simple

Algunas consideraciones

3 Regresiones MIDAS

4 Regresiones UMIDAS

5 Extensiones Modelos MIDAS

Ecuaciones *Bridge* son regresiones **lineales** que asocian variables en alta frecuencia (como el índice de producción industrial, las ventas, la tasa de desempleo) con variables de baja frecuencia (como el PIB).

Advertencia: Esto no pertenece al mundo de modelos macroeconómicos tradicionales, dado que el criterio de inclusión de variables **no** se basa en relaciones causales sino en si: la información es pertinente y mejora el pronóstico de la variable objetivo.

1 Introducción

2 Ecuaciones *Bridge*

Notación

Caso simple

Algunas consideraciones

3 Regresiones MIDAS

4 Regresiones UMIDAS

5 Extensiones Modelos MIDAS

- Sea $t = 1, \dots, T$ el índice de la variable de baja frecuencia y m el número de veces en el que la alta frecuencia aparece en la misma unidad de tiempo t .
- y_t^L denota la variable de baja frecuencia, mientras que una serie de alta frecuencia se denota como $x_{t-j/m}^H$, donde $t - j/m$ es el j^{esimo} periodo de alta frecuencia con $j = 0, \dots, m - 1$.
- Para una mezcla de frecuencias trimestral/mensual, $x_t^H, x_{t-1/3}^H, x_{t-2/3}^H$ son el último, penúltimo y primer mes del trimestre t .

1 Introducción

2 Ecuaciones *Bridge*

Notación

Caso simple

Algunas consideraciones

3 Regresiones MIDAS

4 Regresiones UMIDAS

5 Extensiones Modelos MIDAS

Sea la ecuación que caracteriza a y_t^L :

$$y_t^L = a + bx_t^L + u_t^L \quad (1)$$

con u_t^L igual a un término de error que se supone *i.i.d.*. Los parámetros se pueden estimar por mínimos cuadrados ordinarios y se denotan como \hat{a}_T, \hat{b}_T

Suponga que se desea hallar

$$\hat{y}_{T+1}^L = \hat{a}_T + \hat{b}_T x_{T+1}^L \quad (2)$$

Sin embargo, no se tiene x_{T+1}^L disponible, aunque sí se dispone de $x_{T+1-i/m}^H$.

Por ejemplo, si se tienen valores de x^H para los primeros dos meses del trimestre $T + 1$ ($x_{T+1-2/m}^H, x_{T+1-1/m}^H$) solo hace falta estimar x_{T+1}^H .

Ajustando algún modelo univariado para las observaciones en alta frecuencia, es posible obtener el dato faltante, así:

$$\hat{x}_{T+1|T+1-1/m} = \hat{\phi}(L^{1/m})x_{T+1-1/m}^H \quad (3)$$

con $\hat{\phi}(\cdot)$ igual a un operador de polinomio de rezago para un horizonte de predicción a partir de un modelo obtenido sobre una muestra de tamaño $T_H = T \times m + 2$, dado que se incluyen las dos observaciones del nuevo trimestre a pronosticar y

$$(L^{1/m})x_{T+1-i/m}^H = x_{T+1(i+1)/m}^H.$$

Para pronosticar una observación en la serie de baja frecuencia:

- Se reemplaza el regresor no conocido, x_{T+1}^L con realizaciones parciales de la serie en alta frecuencia.
- Se complementa la serie con estimaciones de las observaciones faltantes, usando la ecuación 3.
- Se pronostica la serie de baja frecuencia.

En otros términos, se tiene que:

$$\hat{y}_{T+1|T+1-(m-i)/m} = \hat{a}_T + \hat{b}_T \left[\sum_{j=1}^i a_{m-j} x_{T+1-(m-j)/m}^H \right] + \hat{b}_T \left[\sum_{h=1}^{i-m} a_{m-(i-h)} \hat{\phi}_h(L^{1/m}) x_{T+1-(m-1)/m}^H \right] \quad (4)$$

1 Introducción

2 Ecuaciones *Bridge*

Notación

Caso simple

Algunas consideraciones

3 Regresiones MIDAS

4 Regresiones UMIDAS

5 Extensiones Modelos MIDAS

- Nótese que en el ejemplo anterior, \hat{a}_T y \hat{b}_T se mantienen constantes, mientras se realiza la actualización de la regresión. En la práctica, esto no tiene por qué ser así: se pueden agregar múltiples regresores o modelos de rezagos distribuidos.
- Usualmente, la combinación de frecuencias es trimestral/mensual.
- La selección de indicadores mensuales se basa en metodologías de selección (*stepwise*, general a específico), criterios de información y evaluación de pronóstico.

- 1 Introducción
- 2 Ecuaciones *Bridge*
- 3 Regresiones MIDAS**
 Parametrización
 Pronóstico
- 4 Regresiones UMIDAS
- 5 Extensiones Modelos MIDAS

- Modelos MIDAS (*Mixed-Data Sampling*) son regresiones de forma reducida con pocos parámetros, donde las variables involucradas se encuentran en diferentes frecuencias.
- La respuesta de la variable de alta frecuencia es modelada con polinomios de rezagos distribuidos altamente parsimoniosos.
- El modelo MIDAS básico, para una sola variable explicativa y un pronóstico h_q pasos adelante, está dado por:

$$y_{t+h}^L = a_h + b_h C(L^{1/m}; \theta_h) x_t^H + \epsilon_{t+h}^L \quad (5)$$

con $C(L^{1/m}; \theta_h) = \sum_{j=0}^{j_{\max}-1} c(j; \theta) L^{j/m}$ y

$$C(1; \theta) = \sum_{j=0}^{j_{\max}-1} c(j; \theta) = 1.$$

- $L^m x_t^H = x_{t-m}^H$.

- 1 Introducción
- 2 Ecuaciones *Bridge*
- 3 Regresiones MIDAS**
 Parametrización
 Pronóstico
- 4 Regresiones UMIDAS
- 5 Extensiones Modelos MIDAS

- La parametrización parsimoniosa de los coeficientes rezagados $c(j; \theta)$ es la característica prominente de MIDAS.
- Existen varias formas funcionales para parametrizar el modelo. La más común es la *Exponential Almon Lag* y se define como:

$$c(k; \theta) = \frac{\exp(\theta_1 k + \dots + \theta_Q k^Q)}{\sum_{k=0}^K \exp(\theta_1 k + \dots + \theta_Q k^Q)} \quad (6)$$

- Esta función puede tomar varias formas con pocos parámetros ($Q = 2$ usualmente).
- Está relacionada con la *smooth polynomial Almon lag function*, usada para reducir multicolinealidad en *Distributed Lags*.

- Una especificación alternativa es la *beta lag*, basada en la función beta y que depende de dos parámetros:

$$c(k; \theta_1, \theta_2) = \frac{f(\frac{k}{K}, \theta_1; \theta_2)}{\sum_{k=0}^K f(\frac{k}{K}, \theta_1; \theta_2)} \quad (7)$$

siendo $f(x, a, b)$ igual a:

$$f(x, a, b) = \frac{x^{a-1}(1-x)^{b-1}\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

$$\text{y } \Gamma(a) = \int_0^\infty e^{-x} x^{a-1} dx.$$

- Un caso específico es $\theta_1 = 1$ y $\theta_2 > 1$. Se logran pesos de pendiente descendiente más flexibles que los patrones exponenciales o geométricos.

Hay otras tres parametrizaciones populares por su flexibilidad:

- 1 Esquema lineal, $c(k; \theta) = \frac{1}{K}$, sin parámetros a estimar en la parte rezagada.
- 2 Un esquema hiperbólico, con

$$c(k; \theta) = \frac{g(\frac{k}{K}, \theta)}{\sum_{k=0}^K g(\frac{k}{K}, \theta)}, \quad g(k, \theta) = \frac{\Gamma(k + \theta)}{\Gamma(k + 1)\Gamma(\theta)}$$

Esta función no suele ser más flexible que la beta.

- 3 Un esquema geométrico, donde $|\theta| \leq 1$ y

$$c(k; \theta) = \frac{\theta^k}{\sum_{k=0}^K \theta^k}$$

- Las anteriores parametrizaciones son flexibles. Para diferentes valores de los parámetros θ , estos pueden caer despacio o rápido; incluso tener una forma abultada.
- Estimar los parámetros a partir de la información automáticamente determina la forma de los pesos (*weights*) y el número de rezagos a incluir en la regresión.
- Un modelo MIDAS puede ser estimado con Mínimos Cuadrados No Lineales (NLS).

- 1 Introducción
- 2 Ecuaciones *Bridge*
- 3 Regresiones MIDAS**
Parametrización
Pronóstico
- 4 Regresiones UMIDAS
- 5 Extensiones Modelos MIDAS

- Se desea predecir la primera observación por fuera de muestra para la variable de baja frecuencia, es decir, $h = 1$. En este caso, el modelo se define como

$$\hat{y}_{T+1|T}^L = \hat{a}_{1,T} + \hat{b}_{1,T} C(L^{1/m}; \hat{\theta}_{1,T}) x_T^H \quad (8)$$

- Si se cuenta con i/m observaciones adicionales, el horizonte de pronóstico h es ahora igual a $h - i/m$ y la Ecuación 8 puede reescribirse como:

$$\hat{y}_{T+h|T+i/m}^L = \hat{a}_{h-i/m,T} + \hat{b}_{h-i/m,T} C(L^{1/m}; \hat{\theta}_{h-i/m,T}) x_{t+i/m}^H$$

- Nótese que todos los parámetros dependen del horizonte de pronóstico.

Para un horizonte de pronóstico h , se obtendrán diferentes estimaciones de los parámetros del modelo, dado que se está proyectando en cada caso con un conjunto de información diferente.

La implicación directa es que la regresión MIDAS *necesita ser reestimada específicamente para cada horizonte de pronóstico*.

Esto suele ser usual en un enfoque de pronóstico directo, como lo hace MIDAS.

- 1 Introducción
- 2 Ecuaciones *Bridge*
- 3 Regresiones MIDAS
- 4 Regresiones UMIDAS**
- 5 Extensiones Modelos MIDAS

- Variación de modelos MIDAS que no depende de polinomios de rezagos distribuidos.
- Suponga que m es pequeño (3, como en la mezcla de frecuencias trimestral/mensual).
- En lugar de estimar $b_h C(L^{1/m}; \theta_h)$ o $C(L^{1/m}; \tilde{\theta}_h)$, se estiman los rezagos de manera independiente - de ahí su acepción de *unrestricted* - lo que supone que la regresión U-MIDAS se escribe como

$$y_{t+h}^L = a_h + \lambda_h y_t^L + c_h^0 x_t^H + c_h^1 x_{t-1/m}^H + c_h^2 x_{t-2/m}^H + \dots + c_h^{m\tilde{K}} x_{t-\tilde{K}}^H + \epsilon_{t+h}^L \quad (9)$$

- En este caso, se estiman, además de los coeficientes a_h y λ_h , $1 + m\tilde{K}$ parámetros adicionales.

- Este modelo U-MIDAS es lineal, por lo que puede estimarse por M.C.O.
- El orden de rezago puede diferir entre variables. Además, se usan criterios de información como AIC, BIC o Hannan-Quinn para escoger el rezago óptimo en cada caso.
- Esta variación de modelos MIDAS puede exhibir proliferación de parámetros. Por ello, modelos U-MIDAS solo funciona para valores pequeños de m .

- Un acercamiento al problema de pronóstico es el usar la técnica de pronóstico directo, el cual puede expresarse como:

$$\hat{y}_{T+m|T}^L = \hat{a}_h + \hat{\lambda}_h y_T^L + \hat{c}_h^0 x_T^H + \hat{c}_h^1 x_{T-1/m}^H + \hat{c}_h^2 x_{T-2/m}^H + \dots + \hat{c}_h^{m\tilde{K}} x_{T-\tilde{K}}^H \quad (10)$$

- En general, se puede usar el mismo acercamiento para pronósticos a distintos horizontes:

$$\bar{y}_{T+h_m|T}^L = \bar{a}_h + \bar{\lambda}_h y_T^L + \bar{c}_h^0 x_T^H + \bar{c}_h^1 x_{T-1/m}^H + \bar{c}_h^2 x_{T-2/m}^H + \dots + \bar{c}_h^{m\tilde{K}} x_{T-\tilde{K}}^H \quad (11)$$

Nótese que en este caso los coeficientes se reestiman según el horizonte de pronóstico.

- 1 Introducción
- 2 Ecuaciones *Bridge*
- 3 Regresiones MIDAS
- 4 Regresiones UMIDAS
- 5 Extensiones Modelos MIDAS**

- La extensión natural de este acercamiento es trabajar con un enfoque multivariado:

$$y_{t+h}^L = a_h + b_h^1 C(L^{1/m}; \theta_h^1) x_{1,t}^H + b_h^2 C(L^{1/m}; \theta_h^2) x_{2,t}^H + \dots + b_h^I C(L^{1/m}; \theta_h^I) x_{I,t}^H + \epsilon_{t+h}^L \quad (12)$$

donde I es el número de series en alta frecuencia.

- Es posible combinar variables explicativas en distintas frecuencias, por ejemplo una serie mensual y una trimestral. Esto, dado que cada variable tiene su estructura de polinomio determinada individualmente.
- En la práctica, adicionar variables explicativas complica la estimación de manera excesiva.
- Como alternativa, se propone trabajar con modelos MIDAS de una sola variable explicativa y agrupar los pronósticos resultantes.

Existen otras extensiones interesantes (solo se mencionarán):

- Modelos *Smooth Transition* MIDAS.
- Modelos MIDAS No Paramétricos.
- Modelos MIDAS asimétricos, no lineales y semi paramétricos.
- Modelos *Markov-Switching* MIDAS.
- Regresiones Cuantílicas.