

Proyecciones de tasas de morosidad: Nota metodológica

Este documento pretende ser una aproximación informal a las técnicas estadísticas usadas para elaborar las proyecciones de las tasas de morosidad. El lector interesado en profundizar en estas técnicas estadísticas puede hacer uso de los textos incluidos en la bibliografía de esta nota. También en este documento se comentarán los datos usados para las proyecciones, así como sus fuentes.

Proyecciones de tasa de morosidad agregada

La tasa de morosidad agregada (*tma*) es, a grosso modo, aquella que corresponde tanto a los hogares como al total de los sectores productivos¹. Dicha tasa no es más que la ratio de los préstamos morosos entre el total de los préstamos concedidos por las entidades de crédito, es decir:

$$tma = \frac{\text{prestamos morosos}}{\text{total prestamos}} * 100 \quad (1)$$

Dichos datos son publicados por el Banco de España alrededor del 16 de cada mes. La frecuencia de tanto *prestamos morosos* como *total prestamos* es mensual, y ambas series comienzan en 1962. Se entienden por préstamos morosos aquellos instrumentos de deuda que tienen importes vencidos por cualquier concepto (principal, intereses...) bien con una antigüedad superior a 3 meses, bien por un importe superior al 25% de la deuda.

El enfoque utilizado consiste en realizar proyecciones de su numerador y denominador por separado, y luego calcular su ratio para obtener la proyección de la tasa de morosidad. Las ventajas de este enfoque son varias. Una de ellas es que existen variables que resultan ser predictores de una de las dos componentes de la ratio pero no de la ratio en su conjunto. Una posible razón es que modelar numerador y denominador por separado da mayor flexibilidad a la hora de escoger las formas funcionales de las relaciones entre la variable a predecir y sus predictores². Otra ventaja de la modelización de numerador y denominador por separado es que, debido a las diferentes propiedades estadísticas de *prestamos morosos* y *total prestamos*, así como la diferente relación con

¹ También se incluyen instituciones sin ánimo de lucro y préstamos sin clasificar, pero el peso de estas dos categorías en el agregado es insignificante.

² Por ejemplo, el paro resulta ser una variable explicativa tanto del numerador como del denominador, pero en uno de los casos aparece en niveles y en el otro en logaritmos.

sus predictores, la metodología econométrica usada para la modelización de una de estas variables puede diferir de la implementada para la otra. En todo caso, como análisis de robustez, se ha probado a predecir directamente la tasa de morosidad, pero los resultados han sido menos satisfactorios.

a) Proyecciones del numerador

Para las proyecciones del numerador se usa el modelo de corrección del error de Engle y Granger (1987). Básicamente, el enfoque de Engle y Granger consiste en dos pasos³:

1. Modelar una relación de equilibrio a largo plazo -llamada cointegración- entre la variable dependiente -la variable a predecir- y un conjunto de variables explicativas. Esto se hace mediante una regresión de las variables explicativas sobre la variable dependiente. Los residuos de dicha regresión son la estimación de la relación de cointegración.
2. Usar dicha relación de equilibrio, retardos de la variable dependiente y otras variables que capturan su comportamiento a corto plazo para predecir los valores futuros de la variable dependiente.

La idea de Engle y Granger es muy intuitiva: si existe una relación de equilibrio a largo plazo entre dos o más variables (p.ej., consumo y renta), toda desviación de dicho equilibrio es necesariamente temporal, es decir, las variables tenderán a volver a su relación de equilibrio (p.ej. la renta y el consumo no pueden llevar sendas completamente divergentes en el largo plazo). Ello mejorará la predicción de la(s) variable(s) respecto a un modelo que sólo tenga en cuenta el comportamiento a corto plazo de éstas.

La aplicación del paso 1 a las proyecciones del numerador se encuentra en una relación de cointegración entre *prestamos morosos*, deudores a la vista (ambas transformadas mediante logaritmo neperiano, log.), paro registrado y euribor a 1 año⁴. Los deudores a la vista son aquellos préstamos que han resultado impagados pero que no entran dentro de la categoría de morosos, en la mayoría de los casos porque aún no han pasado 3 meses desde su vencimiento. Dicha relación se estima mediante una regresión de las variables explicativas -a las que se añade una tendencia lineal-

³ Para una explicación analítica -pero aplicada- de esta metodología véase, p. ej., Enders (1995).

⁴ Aunque paro registrado y deudores a la vista presentan claras pautas estacionales, no han sido desestacionalizadas para el cálculo de la relación de cointegración, pues el test de cointegración mostraba una relación más débil cuando dichas variables eran desestacionalizadas. La fundamentación teórica de este hecho viene explicada en Patterson (2000).

sobre *prestamos morosos* por Ordinary Least Squares (OLS), dado que OLS es superconsistente cuando existe cointegración, y se disponen de más de 150 observaciones de las variables antes mencionadas.

Tras esto se realiza el paso 2, que es obtener las proyecciones de *prestamos morosos* mediante una regresión de toda una serie de predictores sobre la primera diferencia de su logaritmo⁵. Dichos predictores son retardos (de la primera diferencia) de:

- la propia variable dependiente
- el principal medio de los nuevos préstamos hipotecarios
- la inflación anual
- el índice de confianza de los consumidores
- los deudores a la vista (en logs.)
- el importe de los efectos de comercio impagados (en logs.)
- el valor de los inmuebles propiedad de las entidades de crédito (en logs.)
- una variable dummy que refleja el cambio en la contabilidad de los préstamos morosos a partir de Junio del 2005⁶

A éstas variables explicativas se añade un retardo de los residuos de la regresión de cointegración, la cual ha sido antes explicada. Además, todas aquellas variables que presentaban un comportamiento estacional han sido desestacionalizadas mediante el método X12.

b) Proyecciones del denominador

Para las proyecciones del denominador -el total de créditos- se utiliza una metodología distinta, la de Box-Jenkins (1976). Mediante ella se desarrollan los procesos ARIMA (Autorregresive Integrated Moving Average Process) y SARIMA (Seasonal ARIMA). En estos modelos el valor que toma la variable dependiente en un determinado periodo se supone una función de sus valores pasados, así como de un conjunto de perturbaciones no observables presentes y pasadas⁷. Estos modelos, los cuales se estiman mediante el Filtro de Kalman y Maximum Likelihood, pueden ser ampliados mediante la inclusión de variables exógenas o predeterminadas.

⁵ La primera diferencia de una variable $Y(t)$ no es más que una transformación de la serie en la que a cada dato (p.ej. el correspondiente a Marzo de 2005) se le resta su anterior (p.e. el de Febrero de 2005), es decir, $D.Y(t)=Y(t)-Y(t-1)$.

⁶ Antes de Junio del 2005 el BDE contabilizaba el importe de un préstamo moroso como la parte de éste que no había sido pagada 90 días o más después de su vencimiento (p. ej. una mensualidad de una hipoteca). A partir de esa fecha se pasó a contabilizar su importe como toda la parte del préstamo moroso que aún no había sido amortizada.

⁷ Para una introducción a estos modelos véase, p.ej., Harvey (1993)

La variable dependiente del modelo de regresión utilizado es la primera diferencia del logaritmo de *total prestamos*, la cual se puede interpretar como la tasa de crecimiento del crédito total. Dado que dicha serie presenta un claro componente estacional dentro de su estructura autorregresiva, su modelización se realizará mediante un modelo SARIMA, en concreto un SARIMA aditivo⁸.

A dicho modelo se le añaden una serie de variables predeterminadas. Dichas variables son retardos (de la primera diferencia) de:

- el paro registrado (en logs.)
- el euribor a 1 año (en logs.)
- el indicador de sentimiento económico (en logs.)
- el índice de producción industrial (en logs.)
- el número de matriculaciones de vehículos (en logs.)
- el importe de los efectos de comercio impagados (en logs.)
- el valor de los inmuebles propiedad de las entidades de crédito (en logs.)
- el índice general de la Bolsa de Madrid (en logs.)

Las variables que presentan estacionalidad -tales como el paro y el ipi- no han sido previamente desestacionalizadas, dado que el modelo que las usaba desestacionalizadas presentaba peor ajuste dentro y fuera de muestra (Bayesian Information Criterion y Root Mean Squared Error, respectivamente).

Proyecciones de tasa de morosidad del total de sectores productivos

La tasa de morosidad del total de sectores productivos -a la que llamaremos en adelante tasa de morosidad total, *tmt*- es la ratio de los préstamos morosos entre el total de los préstamos concedidos por las entidades de crédito a los sectores productivos, es decir:

$$tmt = \frac{\text{prestamos morosos } tsp}{\text{total prestamos } tsp} * 100 \quad (2)$$

Dichos datos son publicados por el Banco de España alrededor del día 16 de los meses de Marzo, Junio, Septiembre y Diciembre. La frecuencia de tanto *prestamos morosos tsp* como *total*

⁸ En concreto, la tasa de crecimiento del crédito total tiene valores muy por encima de la media en Junio, Julio y Diciembre, y valores muy por debajo de la media en Enero, Febrero y Agosto.

prestamos tsp es trimestral. donde el valor que alcanzan en cada trimestre corresponde al del último mes de dicho trimestre. Es decir, el dato del primer trimestre corresponde a Marzo, el del segundo a Junio, etc. Existen datos de *total prestamos tsp* desde el cuarto trimestre de 1992, mientras que los datos de *prestamos morosos tsp* comienzan mucho más tarde, en el cuarto trimestre de 1998. Esta es otra razón por la que en este caso es útil modelizar numerador y denominador por separado y luego calcular la ratio: la modelización directa de la ratio, dado que sólo está disponible desde 1998, implicaría desechar información potencialmente útil.

a) Proyecciones del numerador

El principal obstáculo de las proyecciones del numerador, *prestamos morosos tsp*, es el escaso número de observaciones -poco más de 40- de las que se dispone para esta variable, así como su frecuencia trimestral. Esto contrasta con la estimación de su variable análoga en la tasa de morosidad agregada, *prestamos morosos*, para la que se disponían de más de 150 observaciones y cuya frecuencia es mensual. Por otro lado, dado que *prestamos morosos tsp* es un subconjunto de *prestamos morosos*, ambas variables están altamente correlacionadas o, visto de otro modo, la segunda capta una gran parte de la varianza de la primera. Por último, los datos sobre *prestamos morosos* son publicados antes que sus análogos sobre *prestamos morosos tsp*; por ejemplo, el dato correspondiente al primer trimestre de 2009 de *prestamos morosos tsp* fue publicado a mediados de Junio, mientras que su análogo sobre *prestamos morosos* -el dato de Marzo lo fue un mes antes, a mediados de Mayo.

Todas estas razones hacen que *prestamos morosos* sea un predictor muy potente de *prestamos morosos tsp*. Es decir, para predecir el dato del primer trimestre de *prestamos morosos tsp* se puede utilizar el dato *observado* de *prestamos morosos* de Marzo, el cual es publicado un mes antes que el primero. Sin embargo, esto implicaría realizar una predicción a muy corto plazo -un mes- y no una proyección a largo plazo, que es el objetivo del presente estudio. Para la proyección de *prestamos morosos tsp* en un determinado trimestre se usa el dato *predicho* de *prestamos morosos* para el mes correspondiente. Por ejemplo, para realizar la proyección del dato de *prestamos morosos tsp* del tercer trimestre se usa la proyección del dato de *prestamos morosos* de Septiembre.

En términos matemáticos:

$$D.\log (\text{prestamos morosos tsp})(t) = a + b * D.\log (\text{prestamos morosos})(t) + \dots + e (t) \quad (3)$$

donde D es la primera diferencia de la variable, $e(t)$ es una perturbación estocástica no observable, y “...” expresa que hay otros regresores en la ecuación.

Sin embargo, el uso de *prestamos morosos* como predictor tiene un problema obvio: es claramente una variable endógena, dado que contiene toda la información presente en la variable dependiente más otra información adicional, y por lo tanto está correlacionada con la perturbación $e(t)$. Dicha endogeneidad lleva a coeficientes sesgados e inconsistentes. No obstante, hay una manera de solucionar este problema, que es recurrir a variables instrumentales⁹.

Las variables instrumentales utilizadas son los propios retardos de *prestamos morosos*. El método de estimación utilizado es Limited Information Maximum Likelihood, dado que diversos estudios de Montecarlo muestran que presenta mejores propiedades en muestras pequeñas que Two-Stage Least Squares, y que Generalized Method of Moments en el presente caso de perturbaciones homocedásticas¹⁰.

El modelo de predicción que se usa es (1), donde el resto de variables explicativas (“...”) son retardos de:

-la primera diferencia de efectos de comercio impagados

-la tasa de concursos (en logs), la cual no es más que la ratio de empresas concursadas en un determinado trimestre sobre el total de empresas en la economía, es decir:

$$tc = \frac{totalempresasconcuradas(t)}{totalempresas(t)} * 1000 \quad (4)$$

b) Proyecciones del denominador

La estrategia para realizar las proyecciones del denominador de la tasa de morosidad del total de sectores productivos, *total prestamos tsp*, es análoga a la usada para el numerador. Es decir, se usan los valores observados y los predichos de *total prestamos* como la principal variable explicativa, corrigiendo por endogeneidad:

$$D.log (totalprestamos(t)) = a + b * D.log (totalprestamos)(t) + \dots + e(t) \quad (4)$$

⁹ Para una introducción al problema de endogeneidad y a su solución mediante variables instrumentales véase, p.ej., Wooldridge (2003).

¹⁰ Véase Poi (2006), Stock, Wright and Yogo (2002) y los artículos citados en ellos.

donde los otros regresores (“...”) son retardos de la primera diferencia de:

- la propia variable dependiente (en logs)
- expectativas de precios de la vivienda (en logs)
- valor de los inmuebles propiedad de las entidades de crédito (en logs)

Datos y fuentes

- Deudores a la vista: Banco de España, Boletín Estadístico capítulo 4
- Efectos de comercio impagados: INE, Estadística de Efectos de Comercio Impagados
- Euribor a 1 año: Banco de España, Estadísticas
- Expectativas de precios de la vivienda: Comisión Europea, Economic and Financial Affairs
- Indicador de sentimiento económico: Comisión Europea, Economic and Financial Affairs
- Índice de confianza de los consumidores: Comisión Europea, Economic and Financial Affairs
- Índice general de la Bolsa de Madrid: INE, Estadísticas Financieras y Monetarias
- Índice de producción industrial: INE
- Inflación anual (medida mediante el IPC): INE
- Inmuebles propiedad de las entidades de crédito: Banco de España, Boletín Estadístico capítulo 4
- Matriculaciones de vehículos: Ministerio de Economía y Hacienda, Dirección General de Análisis Macroeconómico y Económico Internacional
- Paro registrado: Ministerio de Trabajo e Inmigración
- Préstamos totales y préstamos morosos concedidos por entidades de crédito: Banco de España, Boletín Estadístico, capítulo 4
- Principal medio de los nuevos préstamos hipotecarios: INE, Estadísticas Financieras y Monetarias
- Tasa de concursos: calculada mediante datos procedentes de la Estadística de Procedimiento Concursal, la Estadística de Suspensiones de Pagos y Declaraciones de Quiebras y el Directorio Central de Empresas, todas bases de datos del INE

Bibliografia

Box, G. E. P. and G. M. Jenkins (1976): *Time Series Analysis: Forecasting and Control*, revised edition. Holden-Day, San Francisco, C.A.

Enders, Walter (1995): *Applied Econometric Time Series*. Wiley, Canada.

Engle, Robert E. and Clive W. J. Granger (1987): "Cointegration and Error-Correction: Representation, Estimation and Testing." *Econometrica* 55 (March), 251-276

Harvey, Andrew C. (1993): *Time Series Models*, 2nd edition. The MIT Press.

Patterson, Kerry (2000): *An Introduction to Applied Econometrics; a time series approach*. Macmillan.

Poi, B. P. (2006): "Jackknife instrumental variables estimation in Stata". *Stata Journal* 6, 364-376.

Stock, J. H., J. H. Wright and M. Yogo (2002): "A survey of weak instruments and weak identification in generalized method of moments". *Journal of Business and Economic Statistics* 20, 518-529.

Wooldridge, Jeffrey M.(2003): *Introductory Econometrics: A Modern Approach*. Thomson South-Western, Ohio.