

Nombre del artículo o informe

Author Name Affiliation

Resumen El modelamiento de los prepagos es crucial para predecir potenciales pérdidas, debido a la incertidumbre de la duración real de los créditos. En este trabajo se construirá un modelo de regresión con estructura “ARMA” para el ajuste y generación de pronósticos basados principalmente en proyecciones futuras de las variables exógenas de tipo macroeconómicas consideras. Este modelo permitirá en el futuro generar escenarios y evaluar el impacto en la tasa de prepagos de una entidad financiera.

Keywords: pandoc, r markdown, knitr

Introducción

Los créditos hipotecarios son contratos entre dos entidades, generalmente entre una institución financiera y personas naturales o empresas, con la finalidad de permitir a la segunda parte la obtención de una cierta suma de dinero, por ejemplo para el financiamiento de una compra inmobiliaria, automotriz, entre otros bienes de gran valor. Éstos contratos estipulan por una parte un determinado plazo y tasa de interés para la devolución del dinero; por otra parte, se establecen ciertas flexibilidades, entre las cuales destaca la posibilidad del prepagos parcial o total del crédito recibido. Esto tiene ciertas implicancias para los emisores de créditos, ya que el modelamiento de los prepagos es crucial para predecir potenciales pérdidas, debido a la incertidumbre respecto a su duración real. El objetivo de este trabajo es modelar la tasa de prepagos mensual para un cartera de clientes basado en información histórica y variables exógenas que expliquen la evolución temporal de esta tasa, y que sirva como input para generar pronósticos futuros.

Metodología

En la literatura suele referirse a los prepagos mediante un indicador denominado Conditional

Prepayment Rate (CPR_t) que surge de anualizar la tasa de prepagos observada en el mes t , denominada Single Monthly Mortality Rate (SMM_t):

$$SMM_t = \frac{\sum_{i=1}^n \text{Prepago}_{it}}{\sum_{i=1}^n (\text{Saldo}_{it} - \text{Amortizacin}_{it})}$$

donde n representa el número de préstamos vigentes en el mes t , Prepago_{it} el monto prepagado (total o parcial) por cada crédito, Saldo_{it} saldo de deuda al inicio del mes y Amortizacin_{it} la amortización programada dicho período. La anualización de la SMM se obtiene de la siguiente expresión:

$$CPR_t = 1 - (1 - SMM_t)^{12}$$

Las series históricas de prepagos, pueden ser explicadas también por factores exógenos como variables financieras y macro-económicas, pero la presencia de dependencia serial, hace recomendable la utilización de modelos predictivos de series temporales, entre los más populares se encuentra el modelo **SARIMAX** (Seasonal Auto Regressive Integrated Moving Average eXogenous variables por sus siglas en inglés), que permite modelar series de tiempo no estacionarias, con estacionalidad, incorporando a su vez factores externos que pueden estar afectando el comportamiento de la serie. La descripción matemática de un modelo SARIMAX(p, d, q) \times (P, D, Q) $_S$ es como sigue:

$$\varphi(B)\delta(B^S)(1-B)^d(1-B^S)^D(Y_t - X_t\beta) = \theta(B)\vartheta(B^S)\varepsilon_t$$

Donde Y_t es la serie a modelar, B es un operador de rezago $BY_t = Y_{t-1}$, X es una matriz de regresores, β es un vector de coeficientes, y ε_t es una secuencia de errores no correlacionados. Los valores d y D corresponden al número de diferenciaciones necesarias para quitar efectos de tendencia y de estacionalidad, los cuales al considerar regresores son usualmente cero. Los valores p y q están asociados a la dependencia serial observada a la serie ajustada por los regresores. En el caso de detectar un patrón estacional S , se consideran además los

valores P y Q . Los polinomios asociados al modelo son:

$$\varphi(B) = 1 + \varphi_1 B + \dots + \varphi_p B^p \quad (1)$$

$$\delta(B^S) = 1 + \delta_1 B^S + \dots + \delta_p B^{SP} \quad (2)$$

$$\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q \quad (3)$$

$$\vartheta(B^S) = 1 + \vartheta_1 B^S + \dots + \vartheta_Q B^{SQ}$$

El ajuste de un modelo SARIMAX a los datos se realiza considerando varios aspectos relevantes. Por ejemplo, hay que especificar la estacionalidad S ; la diferenciación d , D ; y los órdenes de rezago p , P , q , Q para la serie de tiempo. Usualmente esta especificación se lleva a cabo inspeccionando los gráficos de autocorrelaciones muestrales. Una vez determinados estos valores, se procede al cálculo de los estimadores para todos los coeficientes. Este paso se puede realizar en base a varios métodos de estimación, entre los que podemos mencionar el método de máxima verosimilitud o el método de momentos Yule-Walker. Una descripción detallada de estos y otros métodos de estimación puede encontrarse en los textos Brockwell and Davis (1991) y Palma (2016).

Información

Tomando como referencia lo realizado en Saito (2018), se recopiló información macro-económica desde el Banco Central de Chile¹ y de la Comisión para el Mercado Financiero de Chile (CMF)².

- BCU10: Bonos bullet emitidos por el banco central a plazos de 10 años.
- IMACEC: Índice mensual de actividad económica, ponderación de ciertos indicadores con la participación de distintas actividades económicas en el PIB del año anterior.
- Tasa de Desempleo: Tasa nacional de desocupación entregada por el INE mediante la Encuesta Nacional de Empleo (ENE).

¹<https://www.bcentral.cl>

²<http://www.cmfchile.cl>

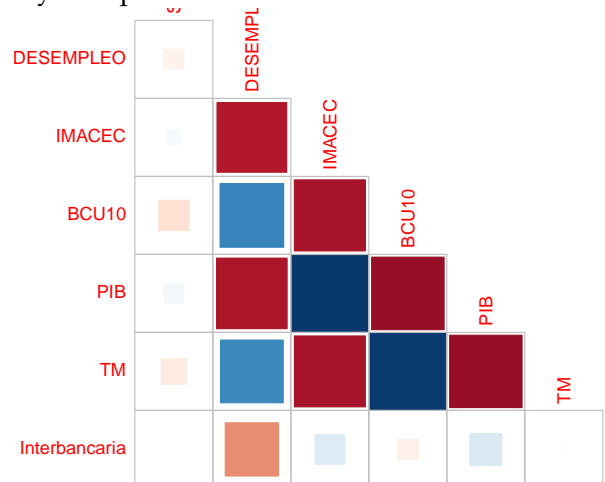
- Tasa de Interés Interbancaria (TIB): Tasa promedio ponderada diaria de préstamos de reservas bancarias entre bancos, sin colateral.
- Tasa de Mercado (TM): Promedio del costo porcentual de captación de los recursos durante el año, más los puntos porcentuales que corresponden a los costos de operación de las instituciones financieras.
- PIB: Producto Interno Bruto, valor total de los bienes y servicios finales (esto es, obtenidos por el consumidor final) producidos por el país, se obtiene a partir de la agregación temporal del IMACEC desestacionalizado.

Con esta información se procedió a modelar la tasa de prepago de créditos hipotecarios en un banco de la plaza.

Modelamiento

Análisis de Correlación

Se realizó un análisis de correlación entre las variables disponibles. La figura ?? se muestra una representación gráfica de la matriz de correlaciones, donde rojo y azul corresponden a -1 y 1 respectivamente.



Nótese en primer lugar que existe poca correlación entre la variable objetivo (SMM) y las demás variables disponibles. En segundo lugar, es necesario poner atención a algunas correlaciones muy potentes entre regresores (como

IMACEC y PIB), lo cual si bien es esperable tratándose de variables macroeconómicas, advierten sobre un posible problema de multicolinealidad entre ellas.

Modelo lineal para SMM

Los posibles modelos lineales construidos se muestran en Tabla ?? . Los modelos 2 – 4 fueron construidos a partir de retirar los regresores no significativos, mientras que el quinto fue construido mediante métodos *stepwise* (cabe mencionar que tanto el método *both* como *backward* entregaron como resultado el mismo modelo). Como se muestra en la misma, el quinto modelo marca mejor ajuste por el criterio de R^2 y un mejor AIC . La columna N.S. indica la cantidad de regresores **no significativos** sobre el total de regresores del modelo, favorablemente el modelo mencionado no posee coeficientes no significativos, por lo que es el modelo más idóneo de entre las opciones presentadas.

##	Modelo	R2	R2aj	AIC	NS
## 1	(1)	0.39336	0.32551	-1533.501	11/19
## 2	(2)	0.38897	0.33378	-1538.276	5/15
## 3	(3)	0.37906	0.32732	-1537.542	5/15
## 4	(4)	0.34989	0.31332	-1537.737	0/10
## 5	(5)	0.38792	0.34531	-1543.983	0/12

Como se muestra en Tabla ?? el quinto modelo marca mejor ajuste por el criterio de R^2 y un mejor AIC . La columna N.S. indica la cantidad de regresores **no significativos** sobre el total de regresores del modelo.

A pesar de lo señalado anteriormente, un análisis de multicolinealidad al modelo seleccionado, arroja que existe un problema de colinealidad. Éste análisis se hará mediante el cálculo del factor de inflación de la varianza VIF por sus siglas en inglés (para mayor información véase [?]), el cual se define concretamente entre TM y BCU10:

```
vif(step01)
```

##	BCU10	PIB	ENE	FEB
----	-------	-----	-----	-----

```
## 11.649147 4.223602 1.159853 1.161208 1.15153
##          OCT          NOV          TM
## 1.149068 1.148345 11.439297
```

Al construir un nuevo modelo retirando el regresor TM y compararlo con los preexistentes, se obtiene que sus métricas son: $R^2 = 0.37384$, $R^2_{aj} = 0.33446$, $AIC = -1542.118$, NS: 0/11. Son cercanas a las del anterior modelo y al aplicar nuevamente el análisis de colinealidad concluimos que está libre de ésta, por lo que finalmente se selecciona éste como modelo final (Ver Código 2)

```
step01
```

```
##
## Call:
## lm(formula = SMM ~ BCU10 + PIB + ENE + FEB + ABR +
##      AGO + OCT + NOV + TM, data = DF)
##
## Coefficients:
## (Intercept)          BCU10          PIB          ENE          FEB
## 8.708e-03      -2.963e-03     -1.474e-07     5.435e-07     3.993e-07
##          MAY          JUL          AGO
## 2.328e-03      3.544e-03      2.356e-03      3.993e-07
```

Modelo SARIMA para los residuos

Se modelarán ahora los residuos del modelo seleccionado con anterioridad como una serie de tiempo. En la Figura ?? se muestran sus gráficos de auto-correlación y auto-correlación parcial, a partir del cual es posible observar su comportamiento estacional, auto-regresivo o de media móvil.

Referencias

Brockwell, P. J., and R. A. Davis. 1991. *Time Series: Theory and Methods*. Second. New York: Springer.

Palma, Wilfredo. 2016. *Time Series Analysis*. John Wiley & Sons.

Saito, Taiyo. 2018. "Mortgage Prepayment Rate Estimation with Machine Learning."

```
ABR      MAY      JUL      AGO
```

PhD thesis, Master's thesis, Delft University of Technology.