

# Proyecciones en tiempo Real del PBI utilizando Indicadores Líderes y elección Bayesiana de variables\*

Fernando J. Pérez Forero<sup>†</sup>

18 de febrero de 2020

## Resumen

There exists a large set of leading indicators that are directly related with GDP growth. However, it is often very difficult to select which of these indicators can be used in order to choose the best short-term forecasting (*nowcasting*) model. In addition, it may be the case that more than one model can do this job accurately. Therefore, it would be convenient to average these potentially non-nested models. Following [Scott y Varian \(2015\)](#), we estimate a Structural State Space model through Gibbs Sampling and a spike-slab prior in order to perform the Stochastic Search Variable Selection (SSVS) method. Posterior simulations can be used to then compute the inclusion probability of each variable for the whole set of models considered. In-sample GDP estimates are very precise, taking into account the large set of regressors considered for the estimation. Data comes from the BCRP's database plus other additional sources.

## Resumen

Existe un gran conjunto de indicadores adelantados que están directamente relacionados con el crecimiento del PBI. Sin embargo, a menudo nos cuesta mucho trabajo seleccionar cuáles de estos indicadores conforman el mejor modelo de regresión y proyección de corto plazo, y que además tenga las características de dar resultados en tiempo real (*nowcasting*). Es más, es posible que exista más de un modelo que pueda realizar esta tarea de manera satisfactoria, por lo que sería conveniente elegir el promedio de estos modelos, los cuales no estarían potencialmente anidados y/o relacionados. Siguiendo a [Scott y Varian \(2015\)](#), se estima un modelo de Estado Espacio mediante el muestreo de Gibbs y un valor *ad hoc* precedente *prior* de tipo *spike-and-slab*, para ejecutar el método de selección estocástica de búsqueda de variables (SSVS). Con ello, se obtiene la probabilidad de inclusión de cada variable dentro del conjunto de modelos posibles. Los resultados muestran un ajuste bastante preciso para el PBI, teniendo en cuenta el gran número de variables utilizadas, las cuales provienen de la base de datos macroeconómicas del BCRP y de otras fuentes adicionales.

JEL Classification: E43, E51, E52, E52, E58

Palabras clave: Proyecciones en tiempo real (*Nowcasting*), Muestreo de Gibbs, Selección de Variables, Promedio de los Modelos.

---

\*Se agradecen los comentarios de los participantes al Encuentro de Economistas XXXVI del BCRP, a los participantes de los seminarios del BCRP. También se agradece los comentarios y sugerencias de dos *referees* anónimos y en especial a Diego Winkelried. Las opiniones expresadas en el presente documento pertenecen al autor y no representan necesariamente las opiniones compartidas por el Banco Central de Reserva de Perú. Todos los errores, fallas u omisiones pertenecen al autor.

<sup>†</sup>Jefe del Departamento del Programa Monetario del Banco Central de Reserva del Perú (BCRP), Jr. Santa Rosa 441, Lima 1, Perú; Correo electrónico: [fernando.perez@bcrp.gob.pe](mailto:fernando.perez@bcrp.gob.pe)

# 1 Introducción

La información es un producto muy valioso para la toma de decisiones. En particular, los responsables de la formulación de políticas y los agentes económicos privados, como por ejemplo los inversionistas, necesitan en todo momento información en tiempo real relacionada con el desempeño de la economía agregada para poder tomar decisiones adecuadas en el futuro. Sin embargo, un tema relevante es el hecho de que la información correspondiente al crecimiento del PBI está disponible con cierto rezago. La duración de este rezago entre países es variable y por lo general este es de más de un mes. Por lo tanto, dado que el PBI actual no es observable, los agentes económicos deben realizar un ejercicio de proyección del presente, es decir, *nowcasting*.

Llevar a cabo esta tarea es posible ya que existe un gran conjunto de indicadores económicos (en particular indicadores líderes) y otras variables que i) están relacionadas con el PBI y ii) se publican casi en tiempo real, es decir, con bastante anticipación al PBI. Como resultado de ello, los economistas (en especial los econométricos), pueden pensar en formular un modelo de regresión lineal para encontrar los determinantes del crecimiento del PBI en base a un subconjunto de estos indicadores líderes (variables adelantadas), y así proyectar el valor actual del PBI. En particular, un ejercicio previo que proyecta el crecimiento del PBI con datos de Perú se puede encontrar en [Pérez-Forero \*et al.\* \(2017\)](#)<sup>1</sup>. Sin embargo, la tarea de proyectar series macroeconómicas en tiempo real no es sencilla, dada la existencia de muchos regresores o determinantes potenciales y especificaciones de modelos que pueden ser planteados, tanto por los diferentes expertos como también por los analistas profesionales expertos en la materia y en econometría. En ese sentido, entre las preguntas que surgen al respecto destacan: ¿cuál es el mejor modelo de regresión lineal a estimar?, ¿cómo podemos seleccionar los mejores regresores entre un gran conjunto de variables?, ¿podemos usar diferentes modelos? En esta línea, es probable que más de un modelo sea popular en algún instante del tiempo, y esto debido a la heterogeneidad de puntos de vista entre los diferentes expertos. De hecho, la mayoría de estos expertos afirman que tienen “el modelo”, y cuando observamos el conjunto de regresores, resulta que estos modelos en muchos casos tienen regresores en común, aunque tienen el inconveniente de no estar anidados. Por ello, en el presente documento se realiza un balance de las opiniones de los diferentes expertos, y con ello se plantea un enfoque ecléctico de proyecciones. Entonces, surge una pregunta más complicada de responder: ¿podemos promediar estos modelos no anidados con diversos y en algunos casos diferentes conjuntos de regresores? Nuestra respuesta es optimista, dado que en este documento se implementa un enfoque econométrico que nos

---

<sup>1</sup>En particular, esta referencia cubre muy bien las investigaciones tempranas relacionadas con el uso de indicadores líderes para el caso peruano (véase, e.g [Escobal y Torres \(2002\)](#), [Ochoa y Lladó \(2003\)](#), y [Kapsoli y Bencich \(2004\)](#), entre otros).

permite obtenerlas.

De este modo, especificamos un modelo estructural de series de tiempo, el cual será estimado mediante técnicas bayesianas siguiendo el enfoque de [Scott y Varian \(2015\)](#)<sup>2</sup>. Dicho modelo viene siendo utilizado para la proyección en tiempo real (*nowcasting*) de series de tiempo usando un conjunto grande de variables, tales como los datos extraídos de Google (*Google Trends data*). Además, para seleccionar entre diferentes modelos y regresores, implementamos el enfoque *spike-and-slab* desarrollado por [Madigan y Raftery \(1994\)](#) y [George y McCulloch \(1997\)](#), para encontrar los mejores predictores para el crecimiento del PBI peruano. Finalmente, dada la distribución posterior estimada para este modelo, ejecutamos un ejercicio de proyección en tiempo real y presentamos la función de densidad de proyección derivada en nuestro análisis.

Los resultados indican que los mejores predictores del crecimiento del PBI peruano son i) la producción de electricidad, ii) el volumen de importación de insumos industriales, iii) el consumo interno de cemento y algunas otras variables financieras. En ese sentido, para aquellos lectores que están familiarizados con los indicadores antes mencionados, este ejercicio es solo una confirmación de lo que siempre han venido pensando. Como se mencionó anteriormente, presentaremos la función de densidad de nuestras proyecciones, en lugar de presentar solo una estimación puntual. Esto último permitirá cuantificar el sustento de la proyección y también la probabilidad de ocurrencia de cualquier punto de proyección en particular, el cual podría potencialmente convertirse en un *fanchart*<sup>3</sup>. Nuestros resultados son robustos, teniendo en cuenta que partimos de un conjunto muy amplio de indicadores económicos, que incluyen entre otros a valores contemporáneos y sus rezagos, dejando que el algoritmo seleccione los regresores principales condicionados a los datos observados y al uso estándar de valores precedentes *priors*. Nuestra expectativa es la de utilizar este algoritmo de forma rutinaria como uno de los modelos satélite que permita la proyección a corto plazo del PBI peruano.

El presente documento está organizado de la siguiente manera: la sección 2 describe el modelo estructural de series de tiempo, la sección 3 describe el procedimiento de estimación, por su parte la sección 4 analiza los resultados principales, y finalmente la sección 5 presenta las conclusiones

---

<sup>2</sup>La literatura asociada a las proyecciones en tiempo real del PBI es amplia y continúa creciendo. Asimismo, las referencias principales sobre este materia y que vale la pena mencionar son [Evans \(2005\)](#), [Giannone et al. \(2008\)](#), y [Banbura et al. \(2013\)](#) entre otros. De otro lado, para los casos específicos de economías pequeñas y abiertas (*Small Open Economies*), podemos mencionar los estudios para Brasil, India y Japón discutidos en [Bragoli et al. \(2015\)](#), [Bragoli y Fosten \(2016\)](#) y [Bragoli \(2017\)](#), respectivamente.

<sup>3</sup>En el análisis de series de tiempo, un *fanchart* es un gráfico de líneas simple para datos pasados, y muestra los rangos para los posibles valores de datos futuros junto con una línea que muestra una estimación central o el valor más probable para los resultados futuros. De esta manera, a medida que las proyecciones se tornan cada vez más inciertas, esto, a medida que avanza el futuro, se observará rangos de proyección que se expanden, creando formas de cuña o abanico distintivas, y de allí proviene el término.

## 2 Un modelo estructural de series de tiempo

### 2.1 Configuración del modelo

Considere la siguiente generalización estocástica del modelo clásico de regresión lineal, el cual incorpora una constante y una tendencia (Scott y Varian, 2015). En particular, la variable dependiente  $y_t$ , representa la tasa de crecimiento anual del PBI, y se determina en base a las siguientes ecuaciones:

$$y_t = \mu_t + \gamma_{t,1} + z_t + v_t, \quad v_t \sim N(0, V) \quad (1)$$

$$\mu_t = \mu_{t-1} + b_{t-1} + w_{1,t}, \quad w_{1,t} \sim N(0, W_1) \quad (2)$$

$$b_t = b_{t-1} + w_{2,t}, \quad w_{2,t} \sim N(0, W_2) \quad (3)$$

$$\gamma_{t,1} = - \sum_{i=1}^{S-1} \gamma_{t-1,i} + w_{3,t}, \quad w_{3,t} \sim N(0, W_3) \quad (4)$$

$$\gamma_{t,i} = \gamma_{t-1,i}, \quad i = 1, \dots, S-1 \quad (5)$$

$$z_t = \sum_{i=1}^K \beta_i x_{i,t} \quad (6)$$

donde  $\mu_t$  representa el término de la tendencia, el cual sigue un comportamiento de tipo paseo aleatorio (*random walk*), además  $b_t$  es un término auxiliar también de tipo paseo aleatorio,  $\gamma_{t,1}$  es el componente estacional, asimismo  $S$  representa la frecuencia de los datos<sup>4</sup> y  $z_t$  el componente exógeno. Como señalamos previamente, este esquema es una generalización de un modelo de regresión lineal, donde dicho caso particular se adecuaría si:  $W_1 = W_2 = W_3 = 0$ . Como resultado, nuestro modelo es lo suficientemente capaz de capturar la variación en el tiempo de los componentes de tendencia y estacional. Lo último es deseable, ya que queremos aumentar la precisión de nuestras proyecciones.

Además, el conjunto de regresores  $x_t$  puede ser potencialmente muy grande, por ello consideramos indicadores líderes tanto en el período contemporáneo  $t$  así como en los valores rezagados  $t - j$ . En ese caso, denotamos  $K = \dim(x_t) = \dim(\beta)$ . Como resultado de lo anterior, tenemos una gran cantidad de modelos candidatos, cada uno considerando solo un subconjunto de estos regresores. Así, en el siguiente apartado explicaremos como se determina cada uno de estos modelos y como podemos adaptar este enfoque en la estimación bayesiana del presente modelo de espacio de estados.

---

<sup>4</sup> $S = 4$  en el caso que se use datos trimestrales y  $S = 12$  cuando se utiliza datos mensuales.

## 2.2 Selección de variables de tipo Spike and Slab

Dada la especificación anteriormente mencionada, ahora es necesario explicar como se implementa el procedimiento de Búsqueda y Selección Estocástica de Variables (SSVS). Para ello, primero consideramos  $\gamma$  como un vector que tiene la misma longitud que la lista de posibles regresores  $x_t$ , y que indica si una variable particular  $x_{i,t}$  está incluida o no en el modelo. Es decir,  $\gamma$  es un vector que tiene la misma longitud que  $\beta$ , esto es,  $K = \dim(\gamma)$ , donde  $\gamma_i = 1$  implica que  $\beta_i \neq 0$  y  $\gamma_i = 0$  implica que  $\beta_i = 0$ , siendo  $\beta_i$  el coeficiente asociado con  $x_{i,t}$ . En otras palabras, cada valor posible del vector binario  $\gamma$  representa un modelo de regresión.

Además, dado el valor de  $\gamma$ ,  $\beta_\gamma$  representa el subconjunto de componentes o elementos de  $\beta$  para las cuales  $\gamma_i = 1$ . Por su parte,  $\sigma^2$  representa la varianza residual del modelo de regresión  $\gamma$ . En este contexto, la distribución a priori o *prior* de tipo *spike* y *slab* para la función de distribución conjunta de  $(\beta, \gamma, \sigma^{-2})$  puede expresarse como:

$$p(\beta, \gamma, \sigma^{-2}) = p(\beta_\gamma | \gamma, \sigma^{-2}) p(\sigma^{-2} | \gamma) p(\gamma)$$

En particular, dado que se tiene una cantidad significativa de ceros en cada modelo  $\gamma$ , el componente *spike* de un *prior* de tipo *spike-and-slab* se refiere a la masa de puntos en cero, para lo cual asumimos una distribución independiente de tipo Bernoulli para cada elemento  $\gamma_i$  con una probabilidad de éxito igual a  $\pi_i$ . En consecuencia, la distribución a priori (*prior*) para  $\gamma$  es un producto de las variables aleatorias Bernoulli:

$$\gamma \sim \prod_i \pi_i^{\gamma_i} (1 - \pi_i)^{1 - \gamma_i} \quad (7)$$

En ese sentido es importante remarcar que, cuando no se cuenta con mucha información detallada, o no está disponible, es conveniente fijar todos los valores  $\pi_i$  igual al valor constante  $\pi$ , es decir, un valor *prior* fijo. De hecho, la probabilidad de inclusión del *prior* común puede obtenerse fácilmente del número esperado de coeficientes distintos de cero para el modelo *promedio*. Esto es, si la expectativa de que los coeficientes de  $k$  pertenecientes a  $K$  sean distintos de cero, entonces estableceremos  $\pi = k/K$  como la media del *prior*.

Por otro lado, el componente *slab* es un prior para los valores de los coeficientes distintos de cero, condicionado a la información de qué los coeficientes son distintos de cero, es decir, condicional a  $\gamma$ . Asimismo, sea  $b$  un vector que representa la mediana del *prior* para los coeficientes de regresión y  $\Omega^{-1}$  una matriz prior de precisión, donde  $\Omega_\gamma^{-1}$  denota las filas y columnas de  $\Omega^{-1}$  para los casos en los

que  $\gamma_i = 1$ . Resulta de lo anterior, un *prior* condicional conjugado en un modelo de regresión lineal Bayesiano estándar como en Zellner (1971) y Koop (2003), donde se observa una distribución normal de los coeficientes:

$$\beta_\gamma \mid \gamma, \sigma^{-2} \sim N \left( b_\gamma, \sigma^2 (\Omega_\gamma^{-1})^{-1} \right)$$

y una distribución Gamma-Inversa para la varianza:

$$\frac{1}{\sigma^2} \sim \Gamma \left( \frac{df}{2}, \frac{ss}{2} \right)$$

Además, en línea con la teoría clásica de regresión lineal,  $X'X/\sigma^2$  representa el componente de información total de Fisher, el mismo que puede ser razonablemente parametrizado como  $\Omega^{-1} = \kappa X'X/T$ . Sin embargo, dado que  $X'X$  es potencialmente incompleto en rango, asumimos que:

$$\Omega^{-1} = \frac{\kappa}{T} (wX'X + (1-w) \text{diag}(X'X)) \quad (8)$$

donde  $0 \leq w \leq 1$ . Bajo este supuesto, es posible obtener siempre una matriz definida positiva  $\Omega^{-1}$ . Es importante destacar que el valor actual de  $w$  es muy cercano a 1, dado que se intentan no distorsionar las propiedades de la muestra de datos, y solo se utiliza este atajo para obtener una simulación numérica estable. Así, la información completa relacionada con los priors se muestra en el siguiente cuadro<sup>5</sup>.

Parámetros	Distribución	Hyper-parámetros
$\alpha_0$	Normal	$N(0_{\dim \alpha \times 1}, I_{\dim \alpha})$
$V$	Gamma-Inversa	$IG\left(\frac{1}{2}, \frac{0.01}{2}\right)$
$W_{i=1,2,3}$	Gamma-Inversa	$IG\left(\frac{1}{2}, \frac{0.01}{2}\right)$
$\beta_\gamma$	Normal	$N\left(0_{\dim \beta_\gamma \times 1}, \sigma^2 (\Omega_\gamma^{-1})^{-1}\right)$
$\sigma^2$	Gamma-Inversa	$IG\left(\frac{1}{2}, \frac{0.01}{2}\right)$
$\gamma$	Spike-slab	$\pi_i = 5/K$

Cuadro 1: Priors para los parámetros del modelo de espacio de estados

<sup>5</sup>Donde  $\kappa = 0.25$ ,  $w = 0.9925$  en  $\Omega^{-1} = \frac{\kappa}{T} (wX'X + (1-w) \text{diag}(X'X))$

## 2.3 Promedio Bayesiano de Modelos

Como se mencionó en un apartado anterior, la inferencia bayesiana con los *priors* de tipo *spike-and-slab* es una forma efectiva de implementar el promedio de modelos de forma bayesiana, ya que esto nos permite explorar el espacio de todos los modelos de regresión disponibles. Como resultado de lo anterior, utilizamos técnicas de tipo Markov-Chain Monte Carlo (MCMC), para así obtener la simulación de la distribución posterior de todo el conjunto de parámetros del modelo.

Además, habiendo simulado la distribución posterior para el espacio de parámetros a través de los métodos MCMC, se tendrá que cada réplica de parámetros de la distribución posterior es posible de ser combinado con los datos disponibles, para así producir una proyección de  $y_{t+1}$  para dicha réplica particular. Asimismo, repetir este proceso varias veces nos da como resultado una estimación de la densidad posterior de las proyecciones  $y_{t+1}$ . Como lo señala [Scott y Varian \(2015\)](#), este último enfoque está motivado por la prueba de [Madigan y Raftery \(1994\)](#), la que permite ponderar un conjunto de modelos en vez de usar un único modelo del conjunto<sup>6</sup>. Como consecuencia de lo anterior, podemos simular completamente la función de densidad de la proyección y luego elegir los percentiles correspondientes para obtener una proyección *promedio*.

## 3 Estimación Bayesiana

El modelo estadístico presentado anteriormente debe estimarse para luego obtener la proyección en tiempo real del PBI. En ese sentido, en esta sección usaremos los métodos de Markov Chain Monte Carlo (MCMC), y así realizar esta tarea. Para ello, primero se debe reescribir el modelo correctamente como un Sistema de Espacio de Estados, y luego explicaremos la rutina de simulación que se usará para la estimación bayesiana.

### 3.1 Representación de Espacio de Estados

El modelo (1)-(2)-(3)-(4)-(5)-(6) puede reescribirse como un Sistema de Espacio de Estados con un componente exógeno y matrices que varían en el tiempo, de modo que:

$$y_t = D_t \alpha_t + Z_t X_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, H_t) \quad (9)$$

$$\alpha_t = A_t \alpha_{t-1} + R_t \eta_t, \quad \eta_t \sim N(0, Q_t) \quad (10)$$

---

<sup>6</sup>Respecto al tema de Promedio bayesiano de modelos, ver también [Fragoso et al. \(2018\)](#), [Hoeting et al. \(1999\)](#) y [Raftery et al. \(1997\)](#).

Asimismo, denotemos  $\psi = (\theta, \alpha^T)$  como el conjunto de parámetros del modelo. Así, la distribución posterior completa para el conjunto de parámetros  $\psi$  sería:

$$p(\psi | y^T) = p(\theta, \alpha^T | y^T) \propto p(\theta) p(\alpha_0) \prod_{t=1}^T p(y_t | \alpha_t, \theta) p(\alpha_t | \alpha_{t-1}, \theta)$$

Nótese que en este caso el modelo es lineal y se encuentra normalmente distribuido condicional al valor de  $\gamma$ . En línea con ello, el cálculo analítico de la distribución posterior estaría disponible condicional al valor de  $\gamma$ . En el siguiente apartado describiremos el algoritmo de estimación.

### 3.2 Rutina de muestreo à la Gibbs

La estimación Bayesiana y la simulación posterior generalmente se implementan a través de los métodos de Markov Chain Monte Carlo (MCMC). Así, en esta sección dividimos el conjunto de parámetros en muchos bloques diferentes, de modo tal que la simulación sea factible, es decir, implementamos una rutina de muestreo de Gibbs como en [Scott y Varian \(2015\)](#). La secuencia del algoritmo es la siguiente:

#### Algoritmo 1

1. Simular  $\{\alpha_t\}_{t=1}^T$  de  $p(\alpha_t | y^T, \psi_{-\alpha_t})$ : [Carter y Kohn \(1994\)](#)

$$\alpha_t | y^T, \psi_{-\alpha_t} \sim N(\bar{\alpha}_{t|T}, \bar{P}_{t|T}), \quad t \leq T \quad (11)$$

2. Simular  $V$  de  $p(V | Y^T, \psi_{-V})$ : Gamma-Inversa
3. Simular  $W_1$  de  $p(W_1 | y^T, \psi_{-W_1})$ : Gamma-Inversa
4. Simular  $W_2$  de  $p(W_2 | y^T, \psi_{-W_2})$ : Gamma-Inversa
5. Simular  $W_3$  de  $p(W_3 | y^T, \psi_{-W_3})$ : Gamma-Inversa
6. Simular  $\beta$  de  $p(\beta | y^T, \psi_{-\beta})$ : Normal
7. Simular  $\sigma^2$  de  $p(\sigma^2 | y^T, \psi_{-\sigma^2})$ : Gamma-Inversa
8. Simular  $\gamma$  de  $p(\gamma | y^T, \psi_{-\gamma})$ : Paso de Metropolis como en [George y McCulloch \(1997\)](#)

Nótese que  $\gamma$  es un vector binario, en el que cada componente tiene una distribución discreta, siendo su valor posterior poco estándar en este contexto. Por lo tanto, para obtener  $p(\gamma | y^T, \psi_{-\gamma})$  como en [George y McCulloch \(1997\)](#), implementamos un paso de tipo Metropolis-Hastings. Así, el paso del algoritmo para esta parte es el siguiente:

## Algoritmo 2

1. Generar un valor candidato de  $\gamma^*$  con distribución de probabilidad  $q(\gamma^{(j)}, \gamma^*)$ .
2. Fijar  $\gamma^{(j+1)} = \gamma^*$  con probabilidad:

$$\alpha^{MH}(\gamma^{(j)}, \gamma^*) = \min \left\{ \frac{q(\gamma^*, \gamma^{(j)})}{q(\gamma^{(j)}, \gamma^*)} \frac{g(\gamma^*)}{g(\gamma^{(j)})}, 1 \right\}$$

En caso contrario  $\gamma^{(j+1)} = \gamma^{(j)}$ .

En particular,  $q(\gamma^{(j)}, \gamma^*)$  es tal que  $\gamma^*$  se genera aleatoriamente cambiando un componente de  $\gamma^{(j)}$ . En consecuencia,  $q(\cdot)$  es una distribución de candidatos simétrica. El uso de distribuciones de candidatos simétricas es bastante estándar en la literatura, y nos brinda la posibilidad de simplificar la transición del Kernel asociada con el paso Metropolis-Hastings.

### 3.3 Configuración de la Estimación

Ejecutamos el muestreo de Gibbs para  $K = 500,000$  casos y descartamos las primeras 100,000 réplicas para minimizar el efecto de los valores iniciales. Además, con el fin de reducir la correlación serial en las réplicas, fijamos un factor de ajuste de 100. Como resultado, obtenemos 4,000 réplicas para ejecutar la inferencia estadística, de tal modo que el ratio de aceptación del paso de metropolis asociado con  $\gamma$  es de alrededor de 0,29.

## 4 Resultados

Luego de simular la distribución posterior de los diferentes bloques de parámetros, analizamos las frecuencias obtenidas y capturamos los mejores predictores en nuestro ranking, los cuales se muestran en el gráfico 1. Se puede apreciar que entre los principales regresores hemos detectado las siguientes variables: producción de electricidad, consumo de cemento y volumen de insumos importados, todos ellos en forma contemporánea ( $t$ ). El ajuste correcto del modelo se muestra en la figura 2, donde el sufijo después de cada etiqueta representa el orden del rezago del regresor empezando en 0.

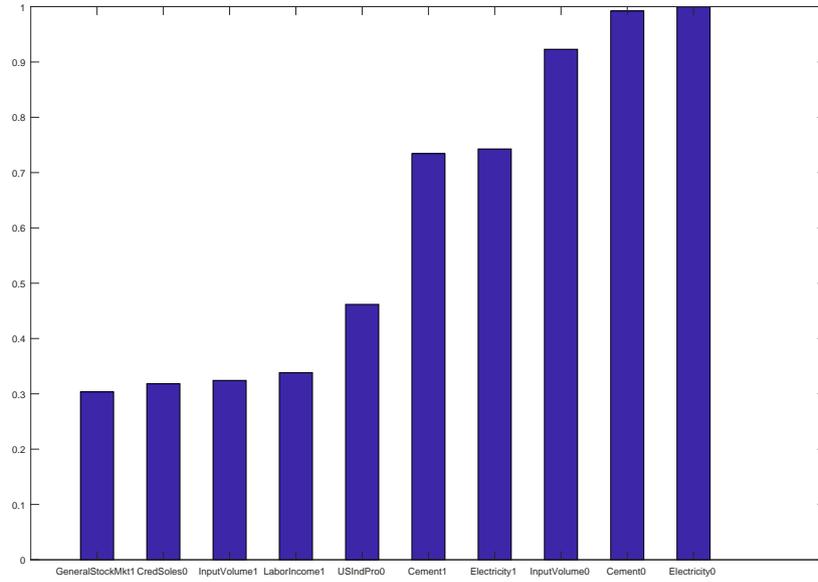


Gráfico 1: Diez principales predictores para el PBI

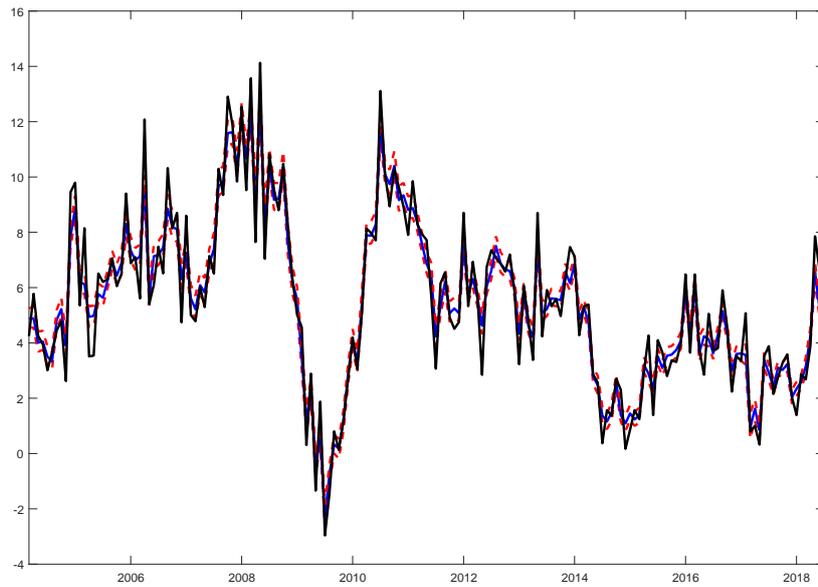


Gráfico 2: Predicción del crecimiento del PBI

Dadas las estimaciones de la distribución a posteriori del conjunto de parámetros  $\psi = (\theta, \alpha^T, \gamma)$ , se tiene que para cada réplica  $i = 1, \dots, S$  de  $\psi$  es posible predecir la variable latente, de modo que:

$$\alpha_{T+h|T}^{(i)} = [A_T^{(i)}]^h \alpha_{T|T}^{(i)} + \eta_{T+h}$$

donde  $\eta_{T+h} \sim N(0, Q_T)$ . Dado esta última estimación, junto con los datos disponibles de los regresores

exógenos hasta un horizonte  $h$ , se podrá proyectar la variable dependiente utilizando la ecuación de medida:

$$y_{T+h|T}^{(i)} = D_{T+h}^{(i)} \alpha_{T+h|T}^{(i)} + (\beta | \gamma)^{(i)} x_{T+h} + v_{T+h}$$

donde  $v_{T+h} \sim N(0, V)$ . Esto es, mientras tengamos datos disponibles del vector  $x$ , será posible calcular la proyección condicional.

## 5 Conclusiones

La proyección del PBI a corto plazo, proceso conocido también como la proyección en tiempo real de esta variable, no es una tarea sencilla. En este trabajo, hemos seleccionado y clasificado los regresores de entre un gran conjunto de variables utilizando técnicas Bayesianas. Entre los principales regresores se ha detectado las siguientes variables: producción de electricidad, consumo de cemento y volumen de insumos importados, todos ellos en forma contemporánea ( $t$ ). No obstante, para nuestro análisis, el método sugerido por [Scott y Varian \(2015\)](#) es muy poderoso y prometedor, siendo el promedio de este modelo el método sugerido. Así, estos resultados nos permiten producir proyecciones de la función de densidad y cuantificar la incertidumbre asociada con la estimación, es decir, el resultado no es solo una proyección puntual del crecimiento del PBI. De allí que además de ser un método muy poderoso resulta ser muy alentador para el hacedor de políticas interesado en analizar diferentes escenarios de riesgo. Sin embargo, se necesita un mayor análisis respecto a la selección *ex-ante* de variables, así como también un análisis de sensibilidad de los resultados utilizando diferentes *priors*. Finalmente, nuestro objetivo también es el de utilizar un conjunto más amplio de indicadores y así aplicar técnicas de *Big Data* como en [Bok et al. \(2017\)](#). En ese sentido, se dejará esto último como parte de la agenda de investigación futura.

## A Descripción de los Datos

Utilizamos datos mensuales desde enero de 2003 hasta julio de 2018 para más de 30 regresores, que incluyen entre otras variables: indicadores de actividad económica, agregados de dinero, indicadores de la BVL, variables del mercado laboral y también variables externas, con valores actuales y rezagados. Así, los detalles se muestran en el cuadro 2.

Cuadro 2: Lista de los regresores incluidos en el modelo

No.	Variable	Unidad de medida original	Transformación	Etiqueta	Fuente
1	Crédito en Soles	Millones de Soles	Variación anual (en %)	CredMN	BCRP
2	Crédito in Dólares	Millones de Dólares	Variación anual (en %)	CredME	BCRP
3	Crédito Total	Millones de Soles	Variación anual (en %)	CredTT	BCRP
4	Crédito Hipotecario	Millones de Soles	Variación anual (en %)	CredHip	BCRP
5	Crédito de Consumo	Millones de Soles	Variación anual (en %)	CredCons	BCRP
6	Depósitos en Soles	Millones de Soles	Variación anual (en %)	LiquMN	BCRP
7	Depósitos en Dólares	Millones de Dólares	Variación anual (en %)	LiquME	BCRP
8	Base Monetaria	Millones de Soles	Variación anual (en %)	Emis	BCRP
9	Efectivo	Millones de Soles	Variación anual (en %)	Circ	BCRP
10	Producción de Electricidad	GWh	Variación anual (en %)	PELEC	COES
11	Consumo de Cemento	Toneladas Métricas	Variación anual (en %)	CINTC	UNACEM
12	Impuestos Indirectos	Millones de Soles	Variación anual (en %)	IGVInt	SUNAT
13	Ventas de Pollo	Prom. diario de Ton. Métricas	Variación anual (en %)	VPOLLOS	BCRP
14	Mano de Obra	Unidades	Variación anual (en %)	PEAO	BCRP
15	Ingreso Laboral Mensual	Millones de Soles	Variación anual (en %)	Ingreso	BCRP
16	Tasa de Desempleo	Porcentajes	Niveles	Desempleo	BCRP
17	Gastos No Financieros del Gobierno	Millones de Soles	Variación anual (en %)	GNF	BCRP
18	Volúmen de Insumos Importados	Unidades	Variación anual (en %)	VOL_M_INPUT	BCRP
19	Índice General de la BVL	(31/12/91 = 100)	Variación anual (en %)	IGBVL	BCRP
20	Índice Selectivo de la BVL	(31/12/91 = 100)	Variación anual (en %)	ISBVL	BCRP
21	IPC	(2009=100)	Variación anual (en %)	IPC	BCRP
22	IPC sin alimentos y energía	(2009=100)	Variación anual (en %)	IPCae	BCRP
23	IPC al por mayor	(2009=100)	Variación anual (en %)	IPM	BCRP
24	Términos de Intercambio	(2007=100)	Variación anual (en %)	TI	BCRP
25	Tasa Libor a 3 meses	Porcentajes	Niveles	Libor3M	FRED
26	Spread del EMBI Peruano	Puntos Básicos	Niveles	EMBI	BCRP
27	Precios del Petróleo (WTI)	Dólares por Barril	Variación anual (en %)	WTI	FRED
28	Tipo de Cambio Real	(2009=100)	Variación anual (en %)	TCRB	BCRP
29	IPC de Estados Unidos	(1982-1984=100)	Variación anual (en %)	IPCUS	FRED
30	Índice de Producción Industrial	(2012=100)	Variación anual (en %)	INDPRO	FRED
31	Índice de Precios al Productor para todos los commodities	(1982=100)	Variación anual (en %)	PPIACO	FRED
32	CBOE Índice de Volatilidad VIX	Unidades	Niveles	VIXCLS	FRED

## B Distribución Posterior de los Hyper-Parámetros

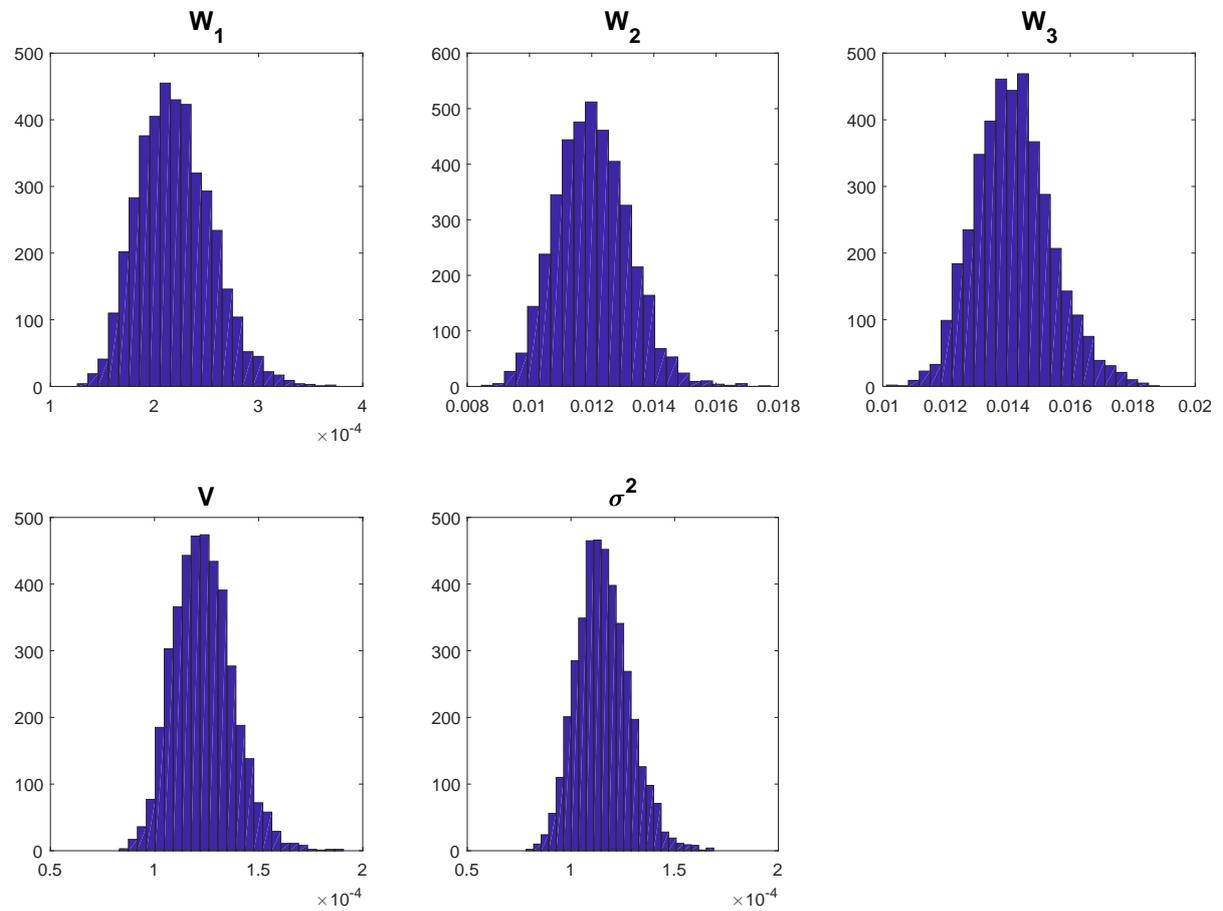


Gráfico 3: Distribución Posterior de los Hyper-Parámetros (Modelo del PBI)

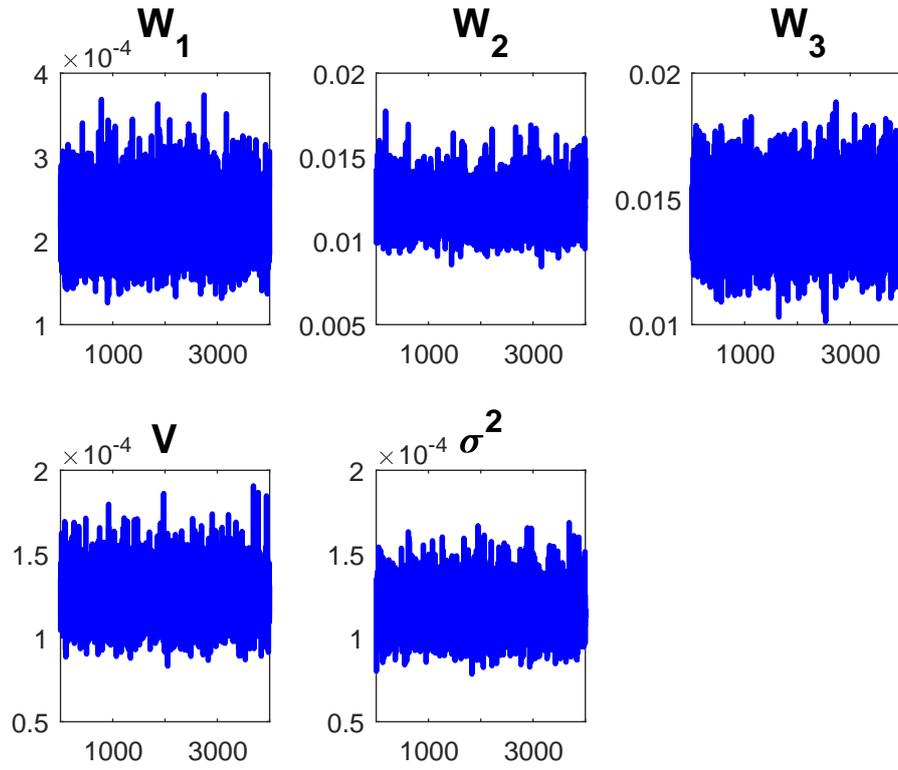


Gráfico 4: Réplicas de la Distribución Posterior de los Hyper-Parámetros

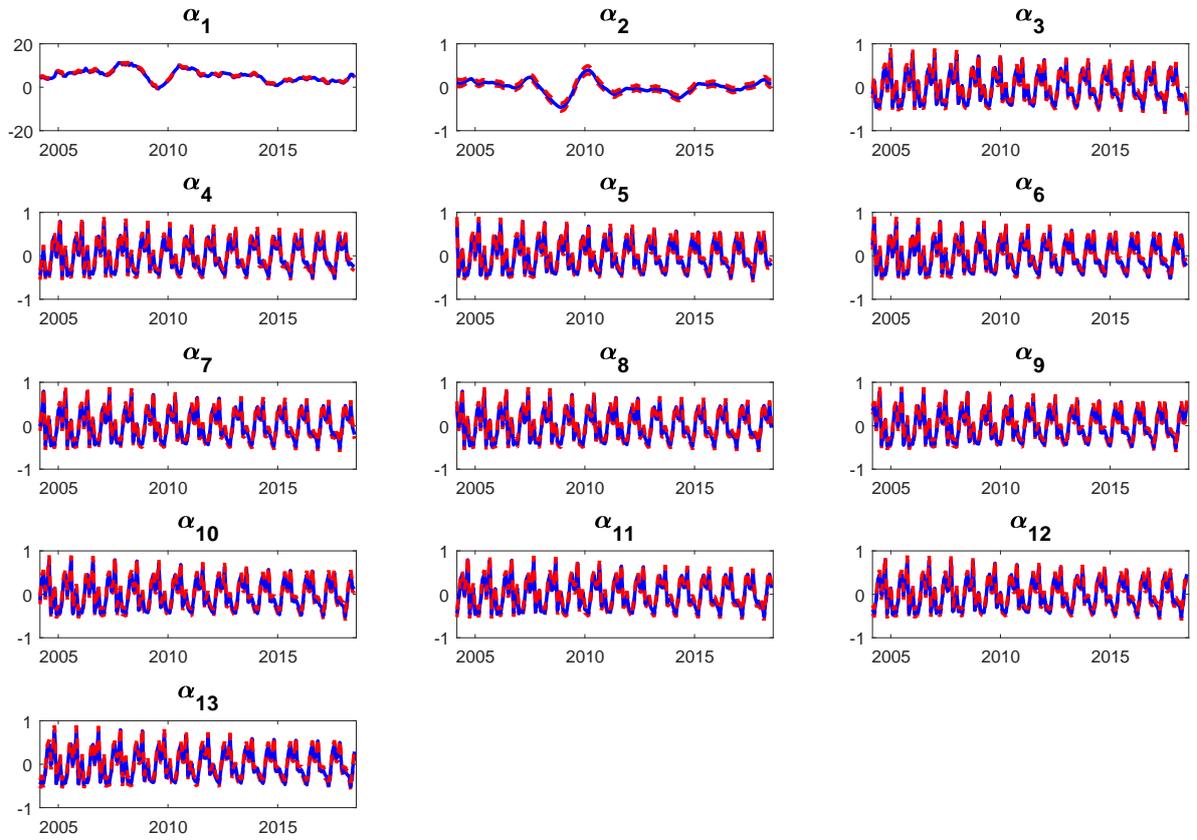


Gráfico 5: Distribución Posterior de los Factores Latentes

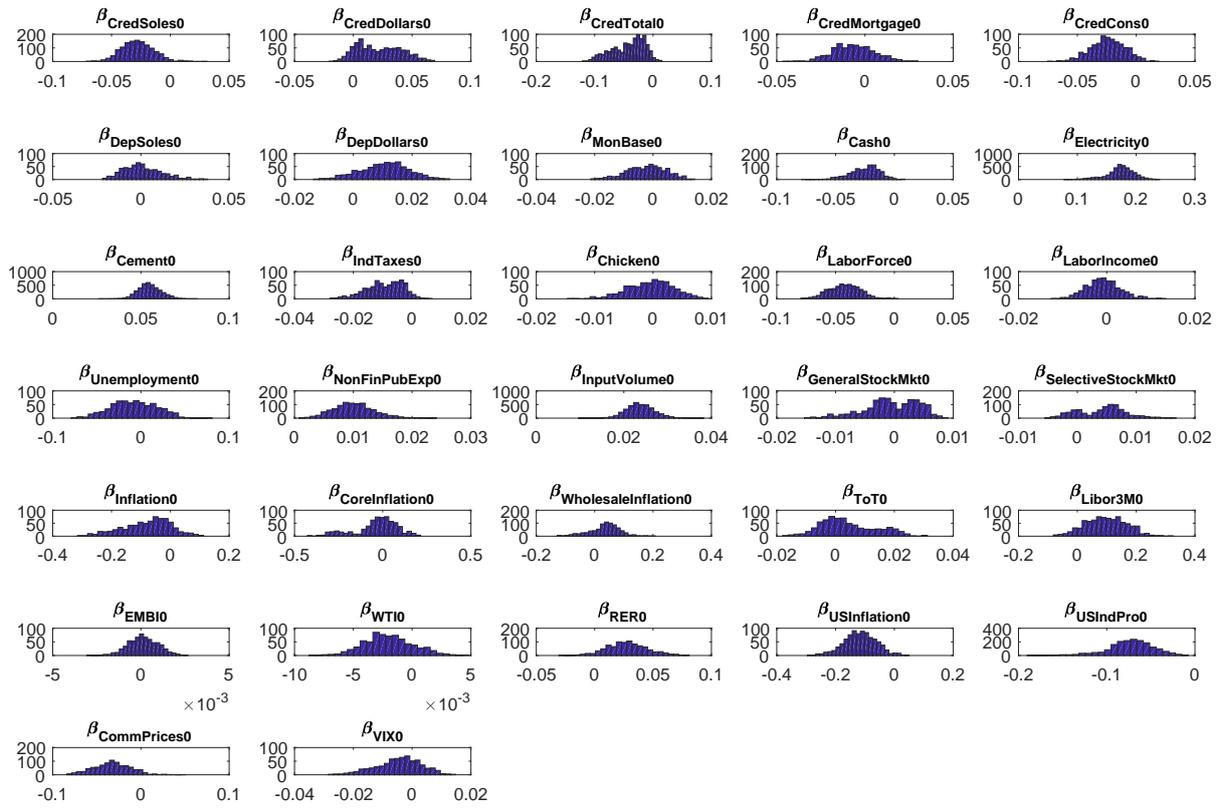


Gráfico 6: Distribución Posterior de los Coeficientes

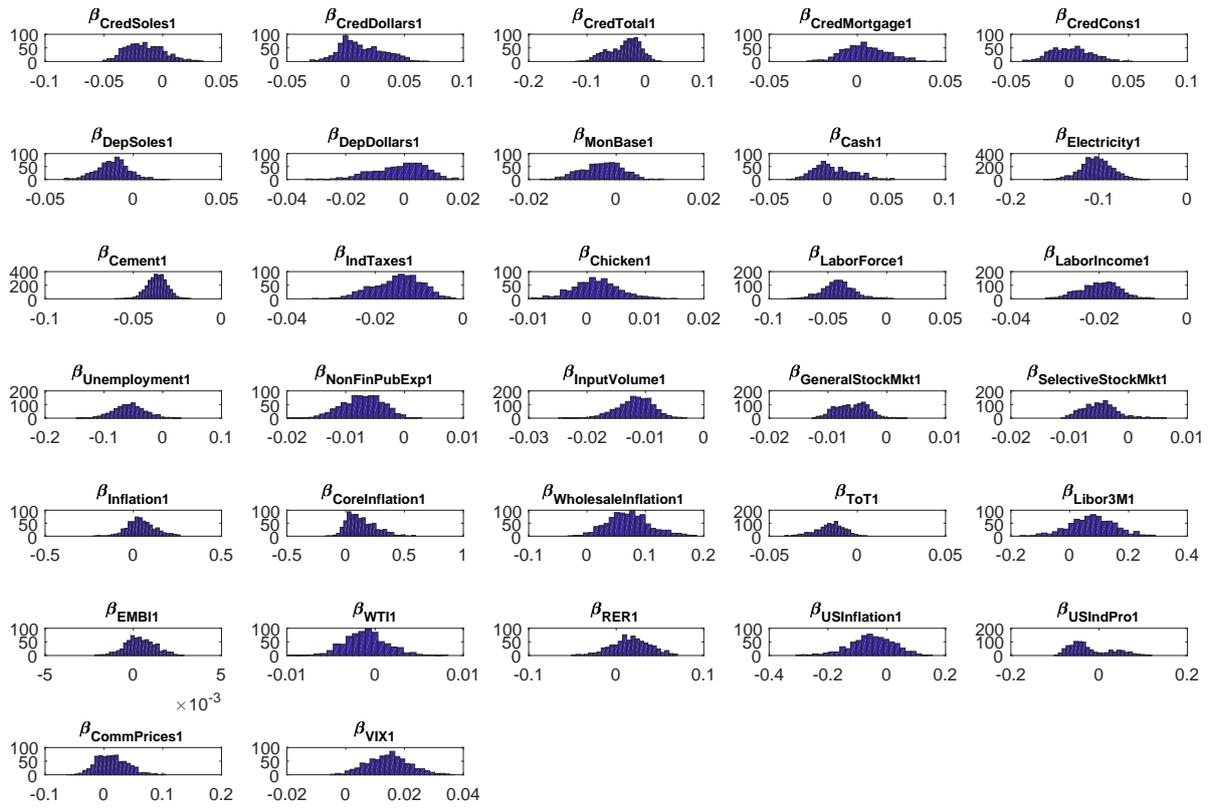


Gráfico 7: Distribución Posterior de los Coeficientes

## Bibliografía

- BANBURA, M., GIANNONE, D., MODUGNO, M. y REICHLIN, L. (2013). *Now-Casting and the Real-Time Data Flow*, Elsevier-North Holland, cap. 2, pp. 2–195.
- BOK, B., CARATELLI, D., GIANNONE, D., SBORDONE, A. y TAMBALOTTI, A. (2017). Macroeconomic nowcasting and forecasting with big data, staff Report No. 830.
- BRAGOLI, D. (2017). Now-casting the japanese economy. *International Journal of Forecasting*, **33** (2), 390–402.
- y FOSTEN, J. (2016). Nowcasting indian gdp, university of East Anglia School of Economics Working Paper Series 2016-06, School of Economics, University of East Anglia, Norwich, UK. June 2016.
- , METELLI, L. y MODUGNO, M. (2015). The importance of updating: Evidence from a brazilian nowcasting model. *OECD Journal: Journal of Business Cycle Measurement and Analysis*, **2015** (1), 5–22.
- CARTER, C. K. y KOHN, R. (1994). On gibbs sampling for state space models. *Biometrika*, **81** (3), 541–553.
- ESCOBAL, J. y TORRES, J. (2002). Un sistema de indicadores líderes del nivel de actividad para la economía peruana.
- EVANS, C. (2005). Where are we now? real-time estimates of the macroeconomy. *International Journal of Central Banking*, **1** (2), 127–175.
- FRAGOSO, T. M., BERTOLI, W. y LOUZADA, F. (2018). Bayesian model averaging: A systematic review and conceptual classification. *International Statistical Review*, **86** (1), 1–28.
- GEORGE, E. I. y MCCULLOCH, R. E. (1997). Approaches for bayesian variable selection. *Statistica Sinica*, **49** (428), 339–373.
- GIANNONE, D., REICHLIN, L. y SMALL, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, **55**, 665–676.
- HOETING, J. A., MADIGAN, D., RAFTERY, A. E. y VOLINSKY, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, **14** (4), 382–401.

- KAPSOLI, J. y BENCICH, B. (2004). Indicadores líderes, redes neuronales y predicción de corto plazo. *Pontificia Universidad Católica del Perú, Revista Economía*, **27**.
- KOOP, G. (2003). *Bayesian Econometrics*. John Wiley and Sons Ltd.
- MADIGAN, D. y RAFTERY, A. (1994). Model selection and accounting for model uncertainty in graphical models using occam's window. *Journal of American Statistical Association*, **82**, 1535–1546.
- OCHOA, E. y LLADÓ, J. (2003). Modelos de indicadores líderes de actividad económica para el Perú. *Revista Estudios Económicos*, **10**.
- PÉREZ-FORERO, F., GHURRA, O. y GRANDEZ, R. (2017). Un indicador líder de actividad real para el Perú, bCRP Documento de Trabajo Documento de Trabajo 2017-001.
- RAFTERY, A. E., MADIGAN, D. y HOETING, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, **92** (437), 179–191.
- SCOTT, S. y VARIAN, H. (2015). Bayesian variable selection for nowcasting economic time series. In A. Goldfarb, S. M. Greenstein y C. E. Tucker (eds.), *Economic Analysis of the Digital Economy*, University of Chicago Press by the National Bureau of Economic Research, pp. 119–135.
- ZELLNER, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. New York, NY: Wiley: reprinted in Wiley Classics Library Edition, 1996.