MINISTERIO DE HACIENDA

SERIE DE DOCUMENTOS DE INVESTIGACIÓN

Modelo FAVAR Óptimo para Proyecciones de Corto Plazo: Aplicación a la República Dominicana

Gustavo Caffaro y Jaime Pérez

No. 2018-01

Ministerio de Hacienda de la República Dominicana

Dirección General de Análisis y Política Fiscal

Modelo FAVAR Óptimo para Proyecciones de Corto Plazo: Aplicación a la República Dominicana

Gustavo Caffaro y Jaime Pérez

Serie de Documentos de Investigación No. 2018-01

Noviembre 2018

Modelo FAVAR Óptimo para Proyecciones de Corto Plazo: Aplicación a la República Dominicana

Gustavo A. Caffaro[1] y Jaime A. Pérez[2]

[1] *Dirección General de Crédito Público, Ministerio de Hacienda*
[2] *Dirección General de Análisis y Política Fiscal, Ministerio de Hacienda*

26 de noviembre de 2018

Resumen

El objetivo de esta investigación es identificar un "mejor modelo" de Factor-Augmented Vector Autoregressive (FAVAR), dadas ciertas restricciones, que proyecte el IMAE y la inflación en la República Dominicana. Para esto se realiza un proceso iterativo en donde se identifican las variables que generan mejor información predictiva. El modelo se compara con un VAR y un FAVAR base, utilizando el criterio Root Mean Squared Error (RMSE). Los resultados indican que el "mejor modelo"FAVAR mejora las proyecciones en más de un 12.9 % para el IMAE y 7.2 % para la inflación, respecto a un modelo base de Vectores Autoregresivos (VAR), y tiene proyecciones 8.9 % y 14.7 % más precisas que el FAVAR base para el IMAE y la inflación, respectivamente.

1. Introducción

El Indicador Mensual de Actividad Económica (IMAE) y la inflación son dos de las variables macroeconómicas más observadas por los hacedores de política y otros agentes económicos. Estos agentes basan gran parte de sus decisiones en el nivel y variación de estas dos variables, formulando un gran número de análisis sobre el impacto de las mismas en sus actividades pasadas, presentes y futuras. La proyección de estas variables es de crucial importancia a la hora de tomar decisiones económicas. Los funcionarios del Banco Central dan fiel seguimiento a las trayectorias del IMAE y la inflación para determinar el nivel de la tasa de política monetaria. Por otro lado, en el Ministerio de Hacienda, los técnicos observan proyecciones de estas dos variables a la hora de determinar la posición del ciclo económico, de aprobar proyectos de inversión, o a la hora de realizar emisiones soberanas, entre otras cosas. Finalmente, los agentes privados formulan conjeturas sobre el IMAE y la inflación para decidir el momento y la plausibilidad de un proyecto de inversión.

En la República Dominicana, existe literatura que intenta realizar proyecciones del crecimiento económico y la inflación (Santana (2017), Jiménez et al. (2015), Paredes (2013)), aunque no existe evidencia de utilizar un modelo FAVAR para dicho fin. El modelo FAVAR cuenta con las bondades de que permite incluir factores que capturan la dinámica de un largo conjunto de variables económicas, superando así uno de los principales problemas de los modelos de series de tiempo. Checo et al. (2017) utilizan este tipo de modelos para medir los efectos de la política monetaria estadounidense en las economías centroamericanas y la República Dominicana. De igual manera, en este trabajo utilizamos la metodología del FAVAR, aunque con el objetivo de realizar proyecciones a corto plazo de la inflación y el IMAE para la República Dominicana.

La metodología central de esta investigación se basa en la de Bernanke et al. (2005), quienes desarrollan un modelo de vectores autoregresivos aumentado con factores (FAVAR, por sus siglas en inglés). Este modelo es una ampliación del modelo de vectores autoregresivos (VAR) donde se consideran, además de las variables objetivo y sus rezagos, una serie de factores, y sus rezagos, que resumen la información de la economía durante distintos períodos. La hipótesis central se basa en que existe un número pequeño de variables inobservables, o factores, que explican la mayoría de los movimientos de las variables observables. Estos factores determinan el comportamiento general de la economía. En este trabajo introducimos un aporte a la literatura, donde seleccionamos el conjunto de variables macroeconómicas que revela la mayor cantidad de información predictiva, explotando las interrelaciones matemáticas entre estas variables. El método utilizado para extraer estos factores es el de componente principal via una descomposición en valores singulares.

Dentro de las principales ventajas de aplicar modelos autorregresivos aumentados con factores es la reducción de la dimensionalidad de los datos. Al extraer los factores, podemos "resumir"la información de las variables, corrigiendo uno de los principales problemas de los VAR: la velocidad con la que se pierden grados de libertad conforme se incluyen nuevas variables. Adicionalmente, este trabajo introduce una metodología de selección de variables, dadas las limitantes al momento de realizar las proyecciones: dentro del amplio espectro de combinaciones posibles de variables, este trabajo intenta encontrar aquel subconjunto de variables que permite realizar las proyecciones más certeras. En otras palabras, en adición a la presentación de la metodología FAVAR en la República Dominicana, este trabajo intenta encontrar un "mejor modelo" dentro de los distintos subconjuntos de variables.

Dentro de las debilidades de este modelo está el hecho de que esta metodología no presta especial atención a la causalidad de las tendencias. Los factores utilizados para realizar las proyecciones resumen la información de la economía de manera tal que resulta difícil realizar inferencias y conclusiones sobre el impacto de las variables exógenas en las variables objetivo. En otras palabras, esta investigación se enfoca no en explicar la relación entre las distintas variables macroeconómicas y su posible comportamiento futuro, sino en encontrar la especificación de un modelo ideal que permita tener proyecciones aceptablemente precisas. En caso de que se quiera realizar un ejercicio de cambio de política, el modelo planteado en esta investigación sirve de poca ayuda. Por lo tanto, el costo de obtener mejores resultados en nuestras proyecciones es perder claridad en las relaciones y las causas de los comportamientos de nuestras variables objetivo. De todas formas, los autores entienden que esta disminuida claridad es justificada cuando el objetivo del ejercicio es obtener las mejores proyecciones disponibles, y más aún cuando es posible realizar investigaciones en paralelo que complementen los resultados de este trabajo.

El resto del documento está distribuido de la siguiente forma: la sección 2 explica el conjunto de variables que conforman nuestra muestra y la transformaciones realizadas; la sección 3 presenta la revisión de la literatura; la sección 4 describe en detalle la metodología utilizada, distinguiendo entre la metodología FAVAR y la propuesta por los autores para la minimización del RMSE; la sección 5 presenta los resultados del mejor modelo FAVAR y se compara con otros modelos tradicionales; y por último, se concluye la investigación.

2. Data

Nuestras variables objetivo son el crecimiento del IMAE y la inflación interanual, medida a través del Índice de Precios al Consumidor (IPC). La muestra utilizada para estimar los factores incluye 53 variables de los cinco sectores de la economía: real, monetario, fiscal, financiero y externo. El espacio temporal recorre los períodos desde el 2008M1 hasta el 2018M6, sumando un total de 126 observaciones. Por tanto, se obtiene una matriz inicial de dimensión 126x53. Es importante destacar que dentro de esta matriz se excluyen las variables objetivo: el crecimiento del IMAE y la inflación.

Para llevar a cabo la extracción de factores de nuestra data por componentes principales, la aplicamos a la misma tres tipos distintos de transformaciones. Primero, siguiendo el procedimiento de Stock y Watson (2002), se normaliza la data para obtener variables con media cero y varianza unitaria. Segundo, se exluyen de nuestro análisis aquellas variables cuya correlación con alguna otra variable dentro de la muestra es mayor a 95 %. La razón es sencilla: el proceso de extracción de componentes principales por valores singulares implica evaluar la correlación dos a dos entre las variables, es decir, calcular la matriz de correlación de nuestra data. Aquellas variables que están altamente correlacionadas entre sí pueden distorsionar el peso atribuido a cada uno de los factores, como por ejemplo el precio de compra del tipo de cambio vs. el precio de venta. Por lo tanto, el movimiento de las variables latentes presentes en la economía ya están debidamente capturados en una de estas dos variables, y el incluir la segunda variable sólo agrega ruido al análisis. Finalmente, se evalua la presencia o no de raíces unitarias en la data, aplicando la prueba de Dickey-Fuller Aumentado a cada variable. En caso de que las variables sean I(1), se toma la primera diferencia interanual. En caso en que no, no se le aplica ninguna transformación adicional.

3. Revisión de Literatura

En la literatura económica los modelos de factores fueron originalmente propuestos por Geweke (1977) y Sargent et al. (1977). Bai y Ng (2002) desarrollaron una teoría inferencial para modelos de factores de largas dimensiones, en la que el estimador de componente principal (PCA, por sus siglas en inglés) fue utilizado por ser fácil de calcular, además de ser asintóticamente equivalente al estimador de máxima verosimilitud. Los modelos FAVAR combinan las ventajas de los modelos de factores con los VAR. Bernanke et al. (2005) representó un punto de partida en el uso de modelos FAVAR para estimar el impacto de shocks en variables macroeconómicas de intereses.

Lanteri (2010) plantea la utilización de FAVAR y de VAR-Bayesiano (BVAR) para pronosticar las exportaciones de Argentina. Se utilizan datos trimestrales desde 1993 hasta el tercer trimestre de 2009. El modelo FAVAR se estima por PCA, mientras que se estiman cinco clases de BVAR. Para evaluar los pronósticos realizan ejercicios dentro y fuera de muestra y calculan el RMSE y el coeficiente de desigualdad de Theil. Los resultados sugieren que los modelos BVAR y FAVAR mejoran las proyecciones realizadas por modelos autorregresivos univariados y VAR si restricciones.

Poghosyan (2013) aplica modelos VAR, VAR-Bayesiano (BVAR) y FAVAR para pronosticar el crecimiento real del PIB, la inflación y la tasa de interés nominal de corto plazo en Armenia. En su trabajo utiliza data trimestral desde el 2000 al 2012. Para comparar la efectividad de los modelos se calcula el RMSE para horizontes desde 1 a 4 trimestres, mediante ejercicios de pronóstico fuera de muestra. Se utilizaron tanto el método recursivo (en donde se va ampliando la muestra) como el de ventana rodante (en el que se mantiene constante el número de la muestra). Los resultados sugieren que en ambos métodos el FAVAR fue mejor para predecir el crecimiento del PIB real.

Figueiredo y Guillén (2013) utilizan un modelo FAVAR para proyectar la inflación de Brasil. Los datos son de frecuencia mensual y van desde junio del 2000 hasta diciembre 2012. Para evaluar la habilidad predictiva del modelo realizan dos métodos de estimación y utilizan tres bases de datos. Para la estimación utilizan el método de dos etapas a través de componentes principales (PCA) y el método de estimación conjunta utilizando cadenas de Markov Monte Carlo (MCMC, por sus cifras en inglés). Para comparar los modelos calcularon los RMSE para horizontes desde 1 a 24 meses, mediante el método de pronóstico recursivo fuera de muestra. Los autores concluyen que el modelo FAVAR supera los AR mientras más largo sea el horizonte de pronóstico. Adicionalmente, el enfoque utilizando MCMC supera los estimados por PCA.

Zardi (2017) utiliza modelos univariales como el Ramdom Walk, SARIMA, el modelo de Time Varying Parameter, BVAR y el FAVAR. Para el pronóstico de la inflación de corto plazo de Túnez. Se utilizan datos trimestrales para el periodo 2000-2015. El FAVAR se estima por PCA, utilizando 30 variables macroeconómicas y financieras. Sus hallazgos revelan que los modelos que incorporan más información económica superan los modelos univariados para proyecciones de horizontes mayores a dos cuatrimestres y que el FAVAR modela bien la data y genera consistentemente un mejor pronóstico.

En cuanto al procedimiento de selección de variables y especificación del modelo, la literatura es bastante amplia, enfocándose más que nada en los algoritmos y procedimientos. Para una breve introducción sobre el procedimiento stepwise y los diferentes criterios de selección de variables, se puede referir a Shumway y Stoffer (2011). Otro procedimiento de selección de variables es la metodología 'lasso', indicada por Tibshirani (1996). Tibshirani no solo propone un método de estimación de modelos lineales distinto al de OLS, sino que este procedimiento también permite realizar selecciones de subconjuntos de variables. Yuan y Lin (2006) amplían este procedimiento a grupos de variables (en vez de variables de manera individual, como Tibshirani), con el objetivo de identificar los factores que mejor predicen la variable objetivo. Un poco más cercano a nuestro dominio de interés, las proyecciones macroeconómicas, Hsu et al. (2008), evalúan diferentes grupos de variables utilizando vectores autoregresivos (VAR) para encontrar el mejor modelo. Finalmente, Gefang (2014) y Koop (2013) desarrollan métodos bayesianos de selección de variables macroeconómicas para proyectar el PIB y la inflación.

4. Metodología

La sección de metodología se divide en dos partes. En la primera parte, se define el modelo FAVAR y se establecen los supuestos. En la segunda parte, se explica el procedimiento de selección del mejor modelo FAVAR. Este proceso se puede dividir en dos etapas: 1) detección de las variables que generan ruido y 2) selección del mejor modelo. Las variables que generan ruido son aquellas variables que, al excluirlas de nuestra muestra, mejoran nuestras predicciones de las variables objetivo. Como punto de comparación, evaluamos el FAVAR base, que contempla el conjunto completo de nuestras variables.

4.1. Especificación del modelo FAVAR

En esta subsección se presenta el modelo FAVAR utilizado para pronosticar nuestras variables objetivo: el crecimiento del IMAE y la inflación. La metodología utilizada se basa en Bernanke et al.

(2005), donde tenemos un vector Y_t de dimensión $m \times 1$ de variables económicas observadas y un vector de factores inobservables que se relacionan con las series de interés. Este vector de factores será F_t de dimensión $K \times 1$ y la dinámica conjunta de ambos vectores viene dado por el siguiente VAR de F_t y Y_t:

$$\begin{bmatrix} F_t \\ Y_t \end{bmatrix} = B(L) \begin{bmatrix} F_{t-1} \\ Y_{t-1} \end{bmatrix} + \epsilon_t \tag{1}$$

Donde $B(L)$ es un polinomio en el operador de rezago, de orden finito d, y ϵ_t es el vector de errores con $E(\epsilon_t) = 0$ y $E(\epsilon_t \epsilon_t') = \Sigma$. El sistema anterior se reduce a un VAR tradicional en Y_t si los términos de $B(L)$ que relacionan Y_t con F_{t-1} son todos ceros. De lo contrario, el sistema 1 será el FAVAR de interés.

La ecuación 1 no puede ser estimada directamente debido a que F_t no es observable. No obstante, asumiendo que existe un vector X_t de dimensión $n \times 1$ de variables informativas tal que $n > K$, su relación con F_t y Y_t es de la siguiente forma:

$$X_t' = \Lambda^f F_t' + \Lambda^y Y_t' + e_t' \tag{2}$$

Donde Λ^f es una matriz $n \times k$ de factores de carga; Λ^y es una matriz de dimensión $n \times m$ y e_t dees $n \times 1$ con media cero y con un pequeño grado de autocorrelación cruzada.

Finalmente, para la estimación del modelo FAVAR se sigue un enfoque de dos pasos de PCA utilizados por Stock y Watson (2002) así como por Bernanke et al. (2005). Para estimar los factores, se realiza la descomposición matricial en valores singulares.

4.2. FAVAR Base

Como fue planteado anteriormente, el argumento principal para utilizar un modelo VAR aumentado en factores es el supuesto de que las variables objetivo dependen, además de sus rezagos, de algunas variables latentes o "señales", en la economía, que son capturadas al observar el comportamiento en conjunto de un gran número de variables macroeconómicas.

El FAVAR base extrae los factores de todas las variables de nuestro dataset, excluyendo, claramente, las variables objetivo. Este modelo nos sirve como punto de comparación para evaluar el desempeño del VAR y del mejor modelo FAVAR. El supuesto del FAVAR base consiste en que todas las variables de nuestra muestra aportan a la obtención de buenas proyecciones del IMAE y la inflación. Este caso no siempre es cierto: como veremos adelante, existen algunas variables, o conjuntos de variables, que, al incluirlas en el FAVAR, nos hacen obtener peores proyecciones de nuestras variables objetivo. A estas variables llamamos "variables que generan ruido".

Una gran desventaja de utilizar un modelo FAVAR es que no es posible utilizar uno de los criterios tradicionales de calidad y parsimonia del modelo, como el Criterio de Información de Akaike o el Criterio de Información Bayesiano (AIC y BIC, respectivamente, por sus siglas en inglés). Como las variables que entran al modelo se resumen en factores, resulta imposible establecer una penalidad por cada variable agregada que cause ruido.

Por esta razón, para comparar el desempeño de los diferentes modelos y a su vez determinar aquellas variables que generan ruido, utilizamos la raiz del error cuadratico medio del pronostico fuera de muestra (out-of-sample RMSE, en inglés), que llamaremos error promedio de predicción, como criterio de evaluación del desempeño de cada modelo. Con este indicador se intenta capturar la distancia entre las predicciones del modelo y los datos observados. Para calcular el out-of-sample RMSE, seleccionamos una fecha o período de corte, t^*, donde se consideran sólo las observaciones anteriores para calcular los estimadores de las regresiones, tal que $t^* < T$ y que 15-40 % de la muestra se encuentre entre t^* y T. [1] A partir de estos estimadores, se proyectan las variables objetivo por un horizonte de h períodos, se almacenan para calcular el RMSE, se aumenta la muestra por un período y se repite el proceso.

Concretamente, estimamos las proyecciones para el período $t+1$ de nuestro vector de variables objetivo, $\hat{Y}_t$:

$$\hat{Y}_{t+1} = X_t \hat{\beta} \tag{3}$$

[1] La literatura está dividida entre cual es el porcentaje de la muestra apropiado para hacer evaluaciones out-of-sample. Nuestra investigación utiliza 30 períodos de 126, resultando en alrededor de 24 % de la muestra dedicado a evaluar el desempeño out-of-sample de los modelos.

donde $X_t = [1 \quad Y_t]$ si incluimos un intercepto, y $\hat{\beta}$ es el vector columna de los coeficientes resultantes de la estimación del modelo 1. Entonces, calculamos el error de predicción, E_t, para cada período t y durante los horizontes $i = 1, 2, ..., h$, de la siguiente manera:

$$E_{t,i} = (\hat{Y}_{t,i} - Y_{t,i}) \tag{4}$$

Cuando $h = 1$, entonces:

$$E_t = (\hat{Y}_t - Y_t) \tag{5}$$

Como nuestra muestra de interés para calcular el desempeño predictivo de nuestro modelo son los períodos a partir de nuestra fecha de corte, t^*, entonces el error promedio de predicción se puede obtener a partir de la siguiente ecuación:

$$RMSE = \sqrt{\frac{\sum_{j=1}^{p} {E_{t+j}}^2}{p}} \tag{6}$$

donde $p = T - t^*$, el número de períodos en los que estamos evaluando nuestro modelo.

4.3. Mejor Modelo FAVAR

Definamos el conjunto de nuestras variables $V = \{v_1, v_2, ..., v_n\}$. Como nuestro objetivo es encontrar el subconjunto de V que, al extraerle los componentes principales, minimiza el error de predicción, nuestro conjunto de interés es el conjunto potencia de V, $P(V)$. Sabemos que si V tiene cardinalidad n, es decir, nuestro número de variables es n, entonces $|P(V)| = 2^n$. Adicionalmente, introduzcamos la notación siguiente: Diremos que si $A \in P(V)$, entonces F_A es el modelo FAVAR que contempla las variables del subconjunto $A \subseteq V$. Por lo tanto, F_V es nuestro modelo FAVAR base, que contempla las n variables en cuestión, y $F_\emptyset$ es el modelo FAVAR que excluye todas las variables (equivalente al modelo VAR).

Como $P(V)$ es un conjunto finito, sabemos que existe un subconjunto de V, que llamaremos, V^*, tal que

$$RMSE(F_{V^*}) \leq RMSE(F_A), \quad \forall A \in P(V) \tag{7}$$

donde $RMSE(\cdot)$ es el RMSE del modelo en cuestión. Es decir, existe un subconjunto V^* que produce el mejor modelo F^* tal que su RMSE es menor o igual al de todos los F obtenidos por subconjuntos de V. En otras palabras, existe una combinación de variables que, al extraer los factores, produce un modelo que realiza las mejores predicciones de las variables objetivo.

Teóricamente, como la cantidad de variables es un número finito, es posible computar los modelos producidos por estas variables. En la práctica, tenemos restricciones, sobre todo de tiempo, computación y de energía, para completar todas las operaciones requeridas para evaluar un número exponencial de modelos diferentes, lo que imposibilita encontrar tal subconjunto F_{V^*}. Una gran complicación es que, al agregar una sola variable a nuestra muestra, duplicaríamos la cantidad de combinaciones posibles de variables. Por ejemplo, si tenemos n variables, entonces necesitamos evaluar 2^n modelos. Si decidimos incluir una variable más, esto aumentaría la cantidad de modelos por evaluar por 2^n.

Como hemos visto, la cantidad de combinaciones posibles crece exponencialmente conforme vamos agregando variables. Quisieramos poder aprovechar la información que nos provee cada variable, pero el costo computacional de añadir una variable se torna cada vez más alto conforme vamos aumentando nuestra muestra. Quisieramos poder disfrutar de los beneficios de añadir nuevas variables, pero sin tener que asumir todos los costos que representan estas nuevas inclusiones.

A pesar de que en la práctica sea imposible evaluar todos los subconjuntos de V, podemos encontrar un subconjunto que nos permita obtener proyecciones suficientemente satisfactorias, dadas nuestras limitantes. Si denotamos K como el número máximo de modelos que podemos evaluar dadas estas restricciones, entonces nos es conveniente encontrar el mayor número entero positivo k tal que $2^k \leq K < 2^{k+1}$. Si tenemos n variables y $k \geq n$, entonces nuestras restricciones no serían limitantes, es decir, seríamos capaces de evaluar todos los modelos posibles resultantes de combinar estas n variables. Regularmente, el caso es que $k < n$, y nos vemos forzados a calcular sólo $2^k < 2^n$ modelos.

Para k muy pequeña o n muy grande, la cantidad de modelos sin evaluar puede ser muy grande. Como argumentamos anteriormente, quisiéramos poder aprovechar la información capturada por todas estas variables y encontrar la combinación que mejor predice nuestras variables objetivo, pero no siempre es factible evaluar todos los modelos posibles para encontrar el mejor de ellos. Para superar este problema, comenzamos por introducir el concepto de ”variable que genera ruido”.

4.4. Variables que Generan Ruido

El objetivo de esta etapa es determinar cuáles son las variables que, al excluirlas de nuestra muestra, ayudan a predecir con mayor exactitud nuestras variables objetivo. En otras palabras, intentamos identificar aquellas variables que, al sacarlas, disminuyen el error promedio de predicción. Para identificar estas variables, y recordando nuestra notación anterior, evaluamos n modelos distintos:

$$F_{V\setminus\{v\}}, \quad \forall v \in V \tag{8}$$

Cada uno de estos modelos FAVAR es construido con todas las variables del dataset menos una. Luego, comparamos el RMSE de cada uno de ellos para determinar los $m \leq n$ modelos con mayor RMSE, es decir, aquellos con peor desempeño.

De esta manera hemos identificado una lista de m variables que queneran ruido. Estas variables son aquellas que, al excluirlas de nuestros modelos, pueden producir mejores resultados predictivos. Si hacemos la relación $m = k$, es decir, igualamos el número m de la lista de variables que generan ruido con el número natural k que restringe el número máximo de modelos por evaluar, entonces podemos asumir que las restantes $(n - k)$ variables son variables "buenas", y nos son útiles para predecir nuestras variables objetivo.

Concretamente, definamos la variable que genera mayor ruido como:

$$\gamma_1 = \Big\{ v \in V \;\big|\; RMSE(F_{V\setminus\{v\}}) \geq RMSE(F_{V\setminus\{u\}}),\ \forall u \in V \Big\}$$

y para $i = 2, 3, ..., k$, la variable que genera el i-ésimo mayor ruido como:

$$\gamma_i = \left\{ v \in \left(V \setminus \bigcup_{j=1}^{i-1} \gamma_j \right) \;\middle|\; RMSE(F_{V\setminus\{v\}}) \geq RMSE(F_{V\setminus\{u\}}),\ \forall u \in \left(V \setminus \bigcup_{j=1}^{i-1} \gamma_j \right) \right\}$$

Esto nos permite definir el conjunto Γ que contiene las k variables que generan ruido:

$$\Gamma = \bigcup_{i=1}^{k} \gamma_i \tag{9}$$

Entonces, el conjunto del resto de las variables "buenas" (que son todas menos las que generan ruido) lo llamamos Φ y está definido por:

$$\Phi = V \setminus \Gamma \tag{10}$$

Ya estamos en una posición en la que podemos determinar la familia de subconjuntos de V que vamos a utilizar para construir la serie de modelos FAVAR. Esta familia X de subconjuntos de V está determinada por:

$$X = \Big\{ \Phi \cup A \;\big|\; \forall A \subseteq \Gamma \Big\} \tag{11}$$

En otras palabras, X es la familia de conjuntos que contiene todas los conjuntos resultantes de las diferentes uniones entre las variables "buenas" y las que generan ruido. Porque $|\Gamma| = k$, entonces $|X| = 2^k$, que es el número máximo de modelos que podemos evaluar dadas nuestras restricciones. Este número es conveniente y deseable porque hemos reducido la cantidad de modelos por evaluar de 2^n modelos a 2^k, un número computable.

De esta manera, podemos encontrar el modelo mejor modelo F_{Λ^*} tal que:

$$F_{\Lambda^*} = \Big\{ \Lambda \in X \;\big|\; RMSE(F_\Lambda) \leq RMSE(F_A),\ \forall A \in X \Big\} \tag{12}$$

Por construcción, F_{Λ^*} es el mejor modelo factible que podemos computar, dadas nuestras restricciones, y sabemos que tendrá igual o mejor poder predictivo que el modelo FAVAR con todas las variables, F_V, y el modelo FAVAR sin ninguna variable (equivalente al VAR), $F_\emptyset$.

En la próxima sección, veremos la aplicación de este procedimiento al caso de la República Dominica.

5. Resultados

En esta sección veremos los resultados de aplicar la metodología del mejor modelo FAVAR al caso de la República Dominicana. Comenzamos comparando los resultados de tres modelos distintos para cada variable objetivo (IMAE e inflación): 1) el modelo VAR con el IMAE y la inflación como variables, 2) el FAVAR base con el IMAE, inflación, y tres factores[2] extraídos de nuestro dataset completo, y 3) el mejor modelo FAVAR con tres factores, obtenido en la ecuación 12, que es el resultante de aplicar la metodología vista en la sección 4. La cantidad de rezagos es de tres para los modelos que optimizan las proyecciones del IMAE y uno para los de inflación, de acorde con el AIC para el VAR.

Para la primera fase de selección del modelo, es decir, para encontrar el conjunto de k variables que generan ruido (ecuación 9), los autores se enfrentaron a una limitante de evaluación de modelos de $K = 2^{15}$. Es decir, seleccionamos las $k = 15$ variables que generan mayor ruido, y luego localizamos el modelo con mejor poder predictivo.

El cuadro 1 presenta los resultados de los errores promedio de proyección. Podemos observar que el error promedio de proyección (RMSE, por sus siglas en inglés) del FAVAR Óptimo fue menor que el del modelo VAR y el modelo FAVAR Base, para las dos variables objetivo: IMAE e inflación. Para el IMAE, el RMSE del FAVAR Óptimo fue de 1.8373, alrededor de 12.87 % y 8.9 % menor que el RMSE del VAR y del Favar Base, respectivamente. Para la inflación, el RMSE del FAVAR Óptimo fue de 0.4121, más de 7.18 % menor que el del VAR y alrededor de un 14.68 % menor que el FAVAR Base.

	RMSE	
	IMAE	Inflación
VAR	2.1087	0.4440
FAVAR Base	2.0163	0.4830
FAVAR Óptimo	1.8373	0.4121

Cuadro 1: Error promedio de proyección para los modelos VAR, FAVAR Base, y FAVAR Óptimo

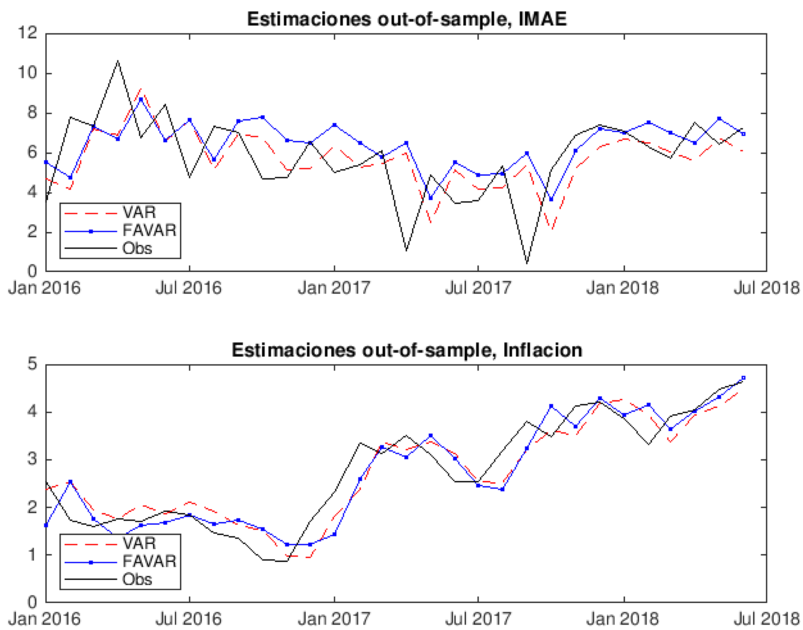

Figura 1: Estimaciones out-of-sample para el IMAE y la inflación

Para tener una idea visual sobre los resultados, la figura 1 grafica las estimaciones out-of-sample de los modelos VAR y el mejor modelo FAVAR para el IMAE y la inflación, en adición a los datos

[2]Estos tres factores son aquellos que explican más del 50 % de la variación total del dataset.

observados, durante los 18 meses del período 2016M1-2018M6. De igual manera, estos resultados se encuentran tabulados en el apéndice.

En cuanto a la distribución de los errores, el cuadro 2 contiene los p-value de los tests de Jarque-Bera y Kolmogorov-Smirnov. Estos tests tienen como hipótesis nula "los errores siguen una distribución normal". Podemos observar que, para el IMAE, no podemos rechazar la hipótesis nula al 92.23 % (Jarque-Bera) ni al 88.17 % (Kolmogorov-Smirnov) de que los errores del FAVAR Óptimo siguen una distribución normal. De igual manera, aplicando las mismas pruebas para la inflación, estas últimas indicaron que no se puede rechazar la hipótesis nula al 58.32 % ni al 92.57 % para los tests de JB y KS, respectivamente.

	IMAE		Inflación	
	Jarque-Bera	Kolmogorov-Smirnov	Jarque-Bera	Kolmogorov-Smirnov
VAR	0.2379	0.7632	0.3861	0.9169
FAVAR Base	0.8142	0.8816	0.5615	0.9951
FAVAR Óptimo	0.9223	0.8817	0.5832	0.9257

Cuadro 2: Test de Jarque-Bera y Kolmogorov-Smirnov, p-value

6. Conclusión

En esta investigación desarrollamos un modelo de Vectores Autorregresivos aumentado en Factores (FAVAR) para la República Dominicana con el fin de proyectar el Indicador Mensual de Actividad Económica y la inflación a corto plazo. Este modelo es el "mejor modelo"FAVAR en el sentido de que, dadas las limitantes de tiempo y poder computacional, presenta mejor desempeño predictivo que un modelo VAR y demás modelos FAVAR. Se encontró que el modelo tiene características estadísticamente deseables, como un bajo error promedio de proyección y normalidad en los errores.

El principal aporte de esta investigación a la literatura es el diseño de una metodología de selección del "mejor modelo"FAVAR utilizando el criterio del RMSE. Esta metodología identifica variables que generan ruido, es decir aumentan el RMSE, mediante un proceso iterativo de exclusión de variables que toma en cuenta las restricciones de tiempo y poder computacional. Adicionalmente, mediante la extracción de factores se reduce la dimensionalidad de la data, el cual es uno de los principales problemas de los modelos autorregresivos.

Referencias

Bai, J. y Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bernanke, B. S., Boivin, J., y Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (favar) approach. *The Quarterly journal of economics*, 120(1):387–422.

Checo, A. M., Pradel, S., Ramírez, F. A., et al. (2017). The effects of usa monetary policy on central america and the dominican republic. *Investigación Conjunta-Joint Research*, pages 189–222.

Figueiredo, F. M. R. y Guillén, O. (2013). Forecasting brazilian consumer inflation with favar models using targeted variables. Technical report, mimeo.

Gefang, D. (2014). Bayesian doubly adaptive elastic-net lasso for var shrinkage. *International Journal of Forecasting*, 30(1):1–11.

Geweke, J. (1977). The dynamic factor analysis of economic time series. *Latent variables in socio-economic models.*

Hsu, N.-J., Hung, H.-L., y Chang, Y.-M. (2008). Subset selection for vector autoregressive processes using lasso. *Computational Statistics & Data Analysis*, 52(7):3645–3657.

Jiménez, M., López, N., y Paredes, E. (2015). Indicador compuesto de crecimiento para la república dominicana. *Documentos de Trabajo del Banco Central de la República Dominicana*, 2015(03).

Koop, G. M. (2013). Forecasting with medium and large bayesian vars. *Journal of Applied Econometrics*, 28(2):177–203.

Lanteri, L. (2010). Modelos de var alternativos para pronósticos (var bayesianos y favar): el caso de las exportaciones argentinas. *Economía*, 33(66):42–64.

Paredes, E. (2013). Indicador adelantado del ciclo económico para la república dominicana. *Documentos de Trabajo del Banco Central de la República Dominicana*, 2013(13).

Poghosyan, K. (2013). Alternative models for forecasting the key macroeconomic variables in armenia.

Santana, L. (2017). Nowcasting con google trends. *Documentos de Trabajo del Banco Central de la República Dominicana*, 2017(03).

Sargent, T. J., Sims, C. A., et al. (1977). Business cycle modeling without pretending to have too much a priori economic theory. *New methods in business cycle research*, 1:145–168.

Shumway, R. H. y Stoffer, D. S. (2011). Time series regression and exploratory data analysis. In *Time series analysis and its applications*, pages 47–82. Springer.

Stock, J. H. y Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460):1167–1179.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.

Yuan, M. y Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67.

Zardi, S. C. (2017). Forecasting inflation in a macroeconomic framework: An application to tunisia. Technical report, Graduate Institute of International and Development Studies Working Paper.

Anexos

A. Tabla de Variables

Nombre de Variable	**Unidad**	**Fuente**	**Sector**
IMAE	Índice	BCRD	Real
Inflación	Índice	BCRD	Precios
Ventas DGII	RD$	DGII	Real
Ventas Agro	RD$	DGII	Real
Ventas Construccion	RD$	DGII	Real
Ventas Explotación de Minas y Canteras	RD$	DGII	Real
Ventas Manufactura	RD$	DGII	Real
Ventas Servicios	RD$	DGII	Real
Ventas comercio otros	RD$	DGII	Real
Ventas comercio combustible	RD$	DGII	Real
Ventas comercio vehículo	RD$	DGII	Real
Ventas comunicaciones	RD$	DGII	Real
Ventas Elec Gas Agua	RD$	DGII	Real
Ventas HB&R	RD$	DGII	Real
Ventas Interm Financiera	RD$	DGII	Real
Ventas Otros Servicios	RD$	DGII	Real
Ventas Transporte y Almacenamiento	RD$	DGII	Real
Spread de Tipo de Cambio Venta	RD$/US$	BCRD	Monetario
Base Monetaria Restringida	RD$	BCRD	Monetario
Operaciones Compra Total de USD	US$	BCRD	Monetario
Encaje Bancario Excedente ME	US$	BCRD	Monetario
Encaje Bancario Excedente MN	RD$	BCRD	Monetario
Medio Circulante (M1)	RD$	BCRD	Monetario
Valores en Circulación	RD$	BCRD	Monetario
Ingresos sin Donaciones	Millones	MH	Fiscal
Gasto Primario	Millones	MH	Fiscal
Gasto Capital	Millones	SIB	Fiscal
Tasa Interbancaria	Porcentaje	BCRD	Financiero
Calidad de Activos	Porcentaje	SIB	Financiero
Tasa de interés activa Promedio Ponderado	Porcentaje	BCRD	Financiero
Tasa de interés pasiva Promedio Ponderado	Porcentaje	BCRD	Financiero
Créditos de OSD al Sector Público MN	RD$	BCRD	Financiero
Créditos de OSD al Sector Privado MN	RD$	BCRD	Financiero
Total Préstamos OSD al Sector Privado	RD$	BCRD	Financiero
Préstamos a la Agricultura, Silvicultura y Pesca	RD$	BCRD	Financiero
Préstamos a Industrias Manufactureras	RD$	BCRD	Financiero
Préstamos a Construcción	RD$	BCRD	Financiero
Préstamos al Comercio Al por Mayor y Al por Menor	RD$	BCRD	Financiero
Préstamos para adquisiciones de Viviendas	RD$	BCRD	Financiero
Tarjetas de Crédito	RD$	BCRD	Financiero
Prestamos de Consumo	RD$	BCRD	Financiero

Nombre de Variable (cont.)	Unidad	Fuente	Sector
Liquidez	Porcentaje	SIB	Financiero
Tasa de interés 3M LIBOR	Porcentaje	FED	Externo
EMBI	Índice	BCRD	Externo
Índice de Producción Industrial	Índice	FED	Externo
Tasa de Desempleo SA US	Porcentaje	BLS	Externo
Tasa de Desempleo US	Porcentaje	BLS	Externo
Tipo de Cambio Real USD/DOP	Indice	SECMCA	Externo
Valor Importaciones petróleo y derivados	US$	BCRD	Externo
Volumen Importaciones petróleo y derivados	Barriles	BCRD	Externo
Llegada de Turistas	Personas	BCRD	Externo
Valor Exportaciones de cacao y manufacturas	US$	BCRD	Externo
Exportaciones de café y manufacturas	US$	BCRD	Externo
Exportaciones de ferroníquel	Tonelada métrica	BCRD	Externo
Exportaciones de azúcar y derivados	Tonelada métrica	BCRD	Externo
Remesas familiares	Millones	BCRD	Externo

B. Tabla de Resultados

	IMAE			Inflación		
	Observado	FAVAR	Δ	Observado	FAVAR	Δ
2016M1	3.5391	6.2259	-2.6868	2.5269	2.3752	0.15177
2016M2	7.8048	4.5785	3.2263	1.7433	2.5342	-0.79096
2016M3	7.3196	5.8646	1.455	1.5886	1.9245	-0.33586
2016M4	10.629	6.7076	3.9211	1.7536	1.5544	0.19922
2016M5	6.766	8.4192	-1.6532	1.7148	1.7665	-0.051687
2016M6	8.4055	7.1986	1.2069	1.9098	1.7053	0.20448
2016M7	4.7515	7.754	-3.0024	1.8548	1.9295	-0.074714
2016M8	7.3215	6.6979	0.62353	1.4724	1.7235	-0.25108
2016M9	7.0194	7.2362	-0.21675	1.3532	1.6176	-0.26443
2016M10	4.667	5.9617	-1.2947	0.91203	1.5504	-0.63833
2016M11	4.7271	6.9038	-2.1767	0.87741	1.0409	-0.16345
2016M12	6.533	6.9453	-0.41234	1.6953	1.1173	0.57803
2017M1	5.0242	6.9201	-1.8959	2.3323	1.8289	0.50342
2017M2	5.4126	5.5889	-0.17634	3.3436	2.5574	0.78623
2017M3	6.1013	5.7264	0.37491	3.1442	3.2909	-0.14663
2017M4	1.0609	5.0609	-3.9999	3.5134	3.0919	0.42153
2017M5	4.8922	4.3013	0.59091	3.1061	3.3916	-0.28559
2017M6	3.4476	4.4155	-0.96793	2.551	3.0029	-0.4519
2017M7	3.5559	2.612	0.944	2.5379	2.4379	0.099999
2017M8	5.321	4.8189	0.50206	3.1825	2.5883	0.59417
2017M9	0.41371	3.4301	-3.0164	3.7992	3.3788	0.42039
2017M10	5.1334	3.1408	1.9926	3.4837	3.9457	-0.46196
2017M11	6.8746	4.8512	2.0234	4.1356	3.6648	0.47071
2017M12	7.3726	5.7443	1.6283	4.2043	4.4265	-0.2222
2018M1	7.0753	7.4111	-0.33573	3.8633	4.4447	-0.58144
2018M2	6.3152	6.0903	0.22486	3.3159	4.0683	-0.75245
2018M3	5.6993	7.3188	-1.6195	3.9113	3.7658	0.1455
2018M4	7.491	6.4883	1.0027	4.0457	4.1574	-0.11169
2018M5	6.4199	6.094	0.32594	4.4704	4.3671	0.10328
2018M6	7.2782	7.1477	0.13055	4.6289	4.745	-0.11612

Cuadro 3: Observaciones, proyecciones out-of-sample, y errores de proyección para los últimos 18 meses de observaciones

C. Gráficos

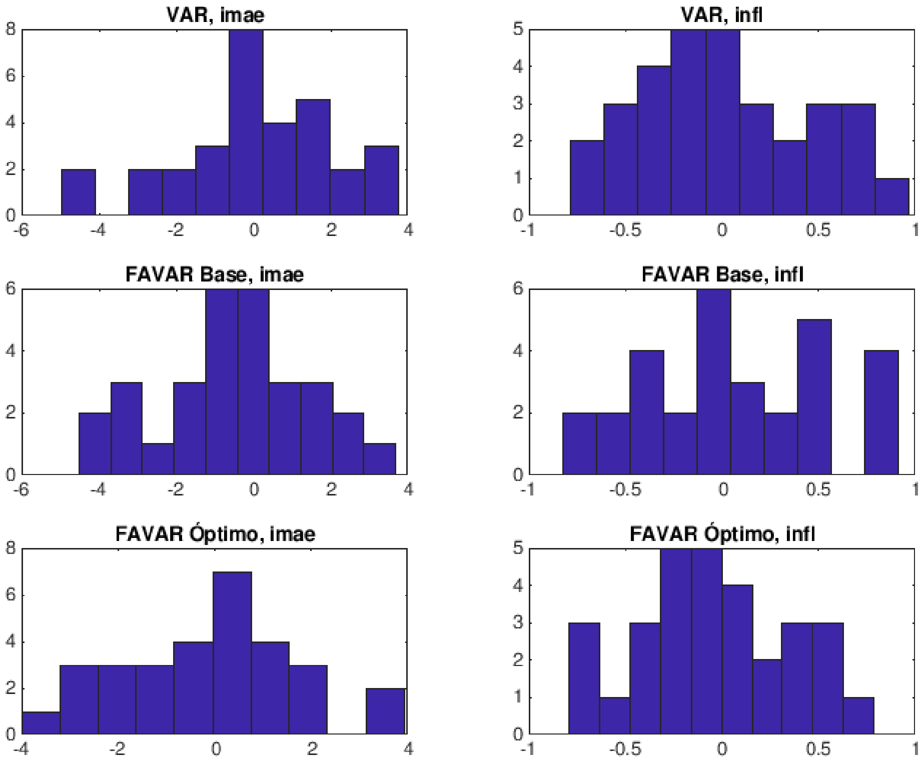

Figura 2: Distribución de los errores para los distintos modelos