

Universidad Torcuato Di Tella – Tesis final de Maestría en Econometría

*Estimación de un indicador subyacente de actividad económica
para Argentina: modelo factorial dinámico*

Abstract

Se presenta la estimación de un indicador mensual de actividad económica para Argentina a partir de la estimación de un factor latente como un promedio ponderado de 32 indicadores mensuales que representan la producción sectorial entre 1995 y 2013. La metodología utilizada consiste en la estimación de un modelo factorial dinámico de tercera generación, extendiendo el modelo factorial estático al campo de las series temporales. Este modelo combina la estimación estática de un factor y su posterior modelización incorporando un componente autorregresivo para capturar la persistencia de las cargas sectoriales de cada indicador. La evaluación de la utilidad de este modelo se concentra en su capacidad predictiva de los ciclos económicos argentinos en tiempo real.

Bárbara Guerezta¹

Alejandra Clemente
Tutora

Junio 2015

¹ *Agradezco los valiosos comentarios de Fernando Marengo, Ricardo Arriazu, Santiago Pérez Vincent, Joaquín Pastor y Francisco Ciocchini, con quienes fui descubriendo la importancia de elaborar esta tesis. Sus comentarios fueron cruciales para estimar un indicador subyacente de la actividad económica que contribuya a nuestra labor como analistas económicos. Gracias por el tiempo y esfuerzo dedicado durante tantas horas de debate.*

Introducción

En Argentina, las cifras trimestrales de Cuentas Nacionales, que muestran la evolución de la actividad económica, se publican con un retraso de 3 meses. Adicionalmente, el INDEC² complementa esta información con la publicación de un anticipo del crecimiento económico mensual a través del “Estimador Mensual de Actividad Económica” (EMAE). Aun así, estos anticipos tienen una demora de entre 45 y 60 días luego de terminado el mes y, adicionalmente, ni el EMAE ni el PBI constituyen indicadores relevantes para el análisis de la situación económica por la falta de credibilidad en las cifras.

Esta realidad demanda un esfuerzo adicional por parte de los analistas, quienes invierten tiempo en construir indicadores para evaluar el pulso de la actividad económica, compitiendo con las cifras oficiales. Coremberg (2014) cuantificó, siguiendo la metodología de Cuentas Nacionales base 2003, el PBI de Argentina entre 1993 y 2012, y encontró evidencia que demuestra la sobreestimación del crecimiento económico por parte de los Organismos de estadísticas oficiales a partir del año 2007. El objetivo de este trabajo es contribuir a esta discusión ofreciendo un indicador para cuantificar la evolución de la actividad económica, pero siguiendo un enfoque estadístico, a partir de un modelo probabilístico, utilizando una muestra de indicadores sectoriales más reducida, publicados por organismos públicos y cámaras de asociación. El indicador que aquí se presenta busca un equilibrio entre celeridad y precisión, sintetizando la información contenida en las series temporales utilizadas, para medir el pulso de la actividad económica, en tiempo real.

Distintas técnicas fueron desarrolladas en los últimos años para reducir la dimensión de la información de la que disponen analistas y hacedores de política, cuya utilidad excede al contexto actual que experimenta Argentina, ya que fueron originalmente diseñadas para obtener una estimación de la actividad que se adelanta, en muchos casos, a la publicación del indicador mensual oficial³. En Argentina, Elosegui, *et.al* (2008) construyó una base de datos con más de 90 indicadores. El número de indicadores sobrepasaba ampliamente la dimensión temporal, lo que le permitió utilizar la metodología de componentes principales estáticos para estimar un indicador real de actividad económica. A diferencia de ése, en este trabajo, la dimensión temporal es superior al número de series utilizadas, por lo que se incluye un análisis de series temporales que

² Instituto Nacional de Estadísticas y Censos.

³ Algunos de los indicadores utilizados en este trabajo se conocen con una demora de más de 60 días luego de terminado el mes. Aunque podría parecer que imposibilita la estimación de un indicador que anticipe la publicación del EMAE, en realidad, se lo puede hacer si se complementa esta técnica con otra que permita la inclusión de paneles desbalanceados para no perder información y obtener estimaciones provisionarias del factor; o bien, con modelos de pronóstico para las series cuya publicación es más tardía, de forma tal de imputar un valor provisorio que permita la estimación del factor.

complementa la estimación en la dimensión transversal, siguiendo la metodología para estimar un modelo factorial dinámico resumida por Stock y Watson (2010).

El Centro de Investigación en Finanzas (CIF) de la Universidad Torcuato Di Tella, emplea una metodología alternativa para estimar en tiempo real la actividad económica. Esta metodología se basa en la utilización de indicadores líderes para construir un indicador suavizado del nivel económico ponderando un conjunto reducido de indicadores seleccionados en base a criterios de conformidad, consistencia temporal, racionalidad económica y representatividad, entre otros.⁴

Los modelos factoriales dinámicos fueron originalmente propuestos por Geweke (1977) como una extensión de series de tiempo de los modelos factoriales previamente desarrollados para datos de corte transversal. En un trabajo anterior, Sargent y Sims (1977) mostraron que dos factores dinámicos podían explicar una larga proporción de la varianza de un conjunto de variables macroeconómicas de Estados Unidos, incluido el PBI, el empleo y los precios.

Stock y Watson (2002) mostraron que el pronóstico de series temporales macroeconómicas utilizando un modelo factorial dinámico superaba a aquellos construidos en base a modelos univariados autorregresivos, vectores autorregresivos y modelos de indicadores líderes.

El trabajo está estructurado de la siguiente forma: en el primer capítulo se presenta el modelo teórico que luego se estimará, siguiendo la tercera generación de modelos factoriales dinámicos; en el segundo capítulo se describen los datos utilizados y las transformaciones realizadas para asegurar la estacionariedad de las series; en el tercer capítulo se presentan los resultados empíricos y se compara el factor estimado con las series de actividad económica que publica frecuentemente el INDEC y en el cuarto capítulo se evalúa la capacidad de predicción del EMAE para el subperíodo 1996-2007 con el factor construido, con el objetivo de validar la técnica utilizada cuando las cifras oficiales no eran cuestionadas, comparando su error de pronóstico con el de un modelo autorregresivo.

⁴ UTDT (2015). Índice Líder. Metodología. Modificación 2015.

1. Modelo Econométrico

El modelo utilizado en este trabajo sigue el diseño de los modelos factoriales dinámicos de la tercera generación⁵, que combinan métodos de estimación de componentes principales y modelos de espacio-estado. La utilización de modelos espacio-estado (cuyos parámetros se estiman a través del filtro de Kalman) como complemento del enfoque estático de componentes principales es útil cuando se quiere modelizar la persistencia en el tiempo de los shocks económicos y las cargas sectoriales en el factor no observable. La esencia de esta especificación radica en la extensión del modelo factorial estático al universo de series temporales. A diferencia de los modelos estáticos, la dimensión temporal es mayor a la transversal, ya que el número de series es inferior al número de observaciones temporales.

En términos generales, siguiendo a Stock y Watson (2010), la premisa de un modelo factorial dinámico es que un conjunto pequeño de factores dinámicos latentes, f_t , dirigen los co-movimientos de un vector de variables económicas temporales estacionarias, X_t , que, a su vez, está también afectado por un vector de errores idiosincráticos con media cero, e_t , que representan los errores de medición y características inherentes a cada uno de los sectores que las variables temporales relevan.

En símbolos, el modelo factorial dinámico puede ser expresado como un modelo estado-espacio,

$$X_t = \beta(L)f_t + e_t \quad (1)$$

$$f_t = \gamma(L) f_{t-1} + \mu_t \quad (2)$$

Donde hay N series; X_t y e_t son de dimensión $N \times 1$; hay q factores dinámicos, por lo que f_t y μ_t son $q \times 1$; L es el operador rezago, y las matrices polinómicas de rezago, $\beta(L)$ y $\gamma(L)$ son $N \times q$ y $q \times q$, respectivamente. La matriz $\beta(L)$ contiene las cargas factoriales dinámicas para las series, que reflejan la contribución del componente común para explicar los co-movimientos de las variables temporales X_t ; Los errores idiosincráticos se asumen incorrelacionados con las innovaciones del factor para todas las cargas y rezagos, esto es: $E e_t \mu'_{t-k} = 0$ para todo k

La ecuación (1) considera sólo las interacciones estáticas entre los indicadores observados a través de su dependencia común sobre el factor latente, f_t . La ecuación (2) permite expandir el modelo para adaptarlo a la dimensión temporal de los indicadores, incorporando la especificación dinámica del factor común y los elementos idiosincráticos.

⁵ Stock y Watson (2010).

La utilización de modelos factoriales dinámicos permite, habiendo estimado los factores f_t y si los errores (e_t, μ_t) son normales, realizar proyecciones eficientes para una variable individual, en función de rezagos de los factores y de la propia variable.

Existen distintos métodos para estimar los parámetros del modelo general que abarca las ecuaciones (1) y (2) – modelos de primera, segunda y tercera generación⁶ -. En este trabajo, el método de estimación que se utiliza sigue a Giannone, Reichlin y Small (2008).

Esencialmente la estimación consta de dos etapas. En primer lugar, el factor latente, f_t , se estima por componentes principales (ecuación -1-) estáticos. En el segundo paso, ese factor estimado, \hat{f}_t , se utiliza para estimar los parámetros desconocidos de la representación de estado-espacio (ecuación -2-).

2. Datos

Esta sección muestra los indicadores utilizados para el análisis y las transformaciones a las que fueron sujetos para asegurar la consistencia de los estimadores.

2.1 Indicadores

Se utilizaron 32 indicadores mensuales⁷, que pueden ser clasificados en dos categorías: 1) los indicadores sectoriales, relevan información de producción de bienes y servicios para cubrir una parte importante de la producción sectorial que releva el PBI argentino; 2) los impulsos (los movimientos de capitales y los términos de intercambio) que contribuyen a explicar los movimientos del PBI desde el enfoque de la demanda, ya que sus movimientos aumentan o contraen el poder de compra interno y generan movimientos del nivel de producción agregada de la economía.

Muchos de los indicadores utilizados sólo están disponibles en términos nominales, por lo cual se los debió deflactar para aislar el impacto que tiene la suba de precios en el nivel de la variable, y de esa forma, obtener un indicador que mida la evolución de la producción en términos reales.

A continuación, se redujo a un número más pequeño de indicadores sectoriales, agrupando aquellos que tenían la misma unidad de medida y representaban al mismo sector (por ejemplo, se

⁶ Stock y Watson (2010).

⁷ Algunas variables (como términos de intercambio y la cuenta capitales), no tienen frecuencia mensual, sino trimestral. Se las mensualizó con E-views para obtener los datos mensuales consistentes con la publicación trimestral que informa el INDEC.

creó una variable agricultura que colapsa la molienda de soja, trigo, maíz, cebada y girasol). Para mayor detalle sobre esta reducción del número de variables se puede consultar el anexo 1. Allí también se presenta el orden de integración de cada variable económica para conseguir la estacionariedad. Luego, se reduce el número de parámetros a estimar – se cuenta con información mensual desde 1994 hasta 2013 –, favoreciendo la parsimonia; y además se evita la sobre-representación de los sectores. La identificación de un factor subyacente está sujeta a la correlación entre las variables que se introducen en el modelo, por lo que la inclusión de muchas variables que representan a un único sector puede distorsionar los resultados, haciendo que el factor refleje más que proporcionalmente los co-movimientos de estas variables.

El gráfico 1 muestra la evolución de todos los indicadores utilizados entre 1995 y 2013. Es importante notar que, otra característica distintiva de este trabajo respecto de otros presentes en la literatura⁸ es que aquí no se incluyó dentro del set de variables utilizadas el Estimador Mensual de Actividad Económica que publica el INDEC (EMAE), sino que se utilizaron indicadores públicos, que en muchos casos coinciden con los que utiliza el INDEC para estimar el EMAE, pero no se utilizó el EMAE para delimitar las ponderaciones básicas dentro del factor.

El gráfico 1 muestra que en la mayoría de los indicadores se observa una caída de la producción en el año 2002 (signada por la crisis cambiaria de la salida de la convertibilidad), su posterior recuperación casi ininterrumpida hasta el año 2007, seguido de un debilitamiento de los índices de producción en el año 2009 y un virtual estancamiento de estos indicadores a partir del año 2012.

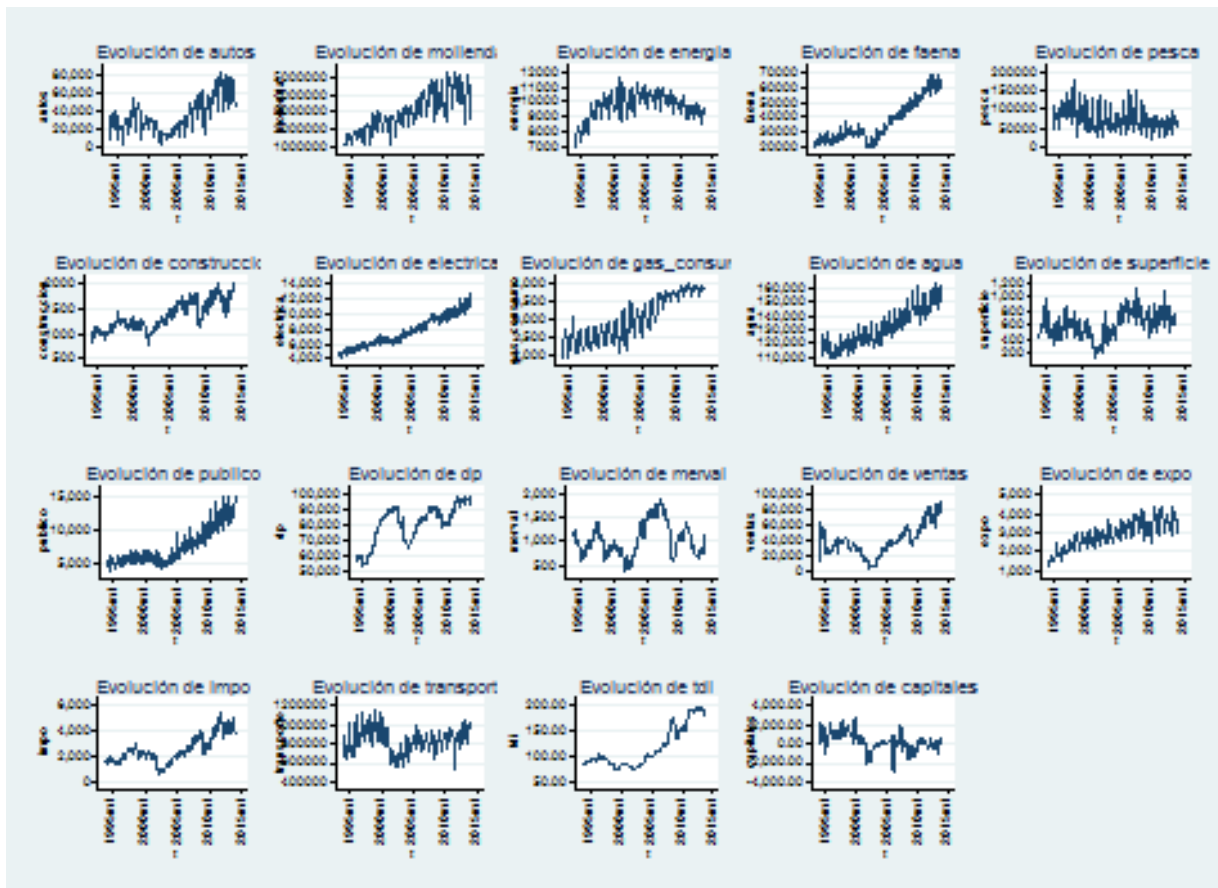
Una clara excepción a esa dinámica la constituye el indicador de energía, que unifica la producción de petróleo crudo, procesado y natural, que luego de alcanzar un máximo a fines de la década de 1990, redujo su producción hasta fines del año 2013.

Una mención especial merece la serie de los términos de intercambio (tdi), cuyo valor se mantuvo estable hasta entrado el año 2003, momento a partir del cual la suba de los precios internacionales de materias primas impulsó el crecimiento de los términos de intercambio, llevándolo a niveles récord históricos. Esta suba generalizada de las materias primas se lo conoce como “viento de cola” para todos los países productores de materias primas, entre ellos, Argentina.

Puede verse que la mayoría de las series presentan una clara tendencia, que exige su extracción para asegurar estacionariedad. El siguiente capítulo explica cómo se extrajeron las tendencias y de qué manera se homogeneizaron las unidades de medida para estimar el modelo.

⁸ Ver Elosegui, et.al (2008).

GRÁFICO 1. INDICADORES SELECCIONADOS



La tabla 1 muestra los indicadores que se utilizaron, el sector que representan, la fuente de información y el rezago con el que se conoce cada dato.

TABLA 1. INDICADORES SELECCIONADOS

Nombre	Sector que Representa	Descripción	Periodo	Fuente	Transformación	Cuántos días pasan hasta que se conoce el dato del mes
autos	Industria	Prod. de Automóviles (UNIDADES)	Mensual 1996-2013	ADEFA	--	5
		Prod. de Automóviles (UNIDADES)	Mensual 1994-1995	MECON Info económica al día	--	5
soja	Agricultura / Industria	Molienda de granos Soja (Tn)	Mensual 1990-2013	MinAgri	--	15-30
trigo	Agricultura / Industria	Molienda de granos Trigo pan (Tn)	Mensual 1990-2013	MinAgri	--	15-30
maíz	Agricultura / Industria	Molienda de granos Maíz (Tn)	Mensual 1990-2013	MinAgri	--	15-30
girasol	Agricultura / Industria	Molienda de granos Maíz (Tn)	Mensual 1990-2013	MinAgri	--	15-30
cebada	Agricultura / Industria	Molienda de Cebada forrajera y cervecera (Tn)	Mensual 1990-2013	MinAgri	--	15-30
crudo	Industria	Producción de Petróleo Crudo (m3)	Mensual 1994-2013	MECON Info económica al día	--	15-30
procesado	Industria	Producción de Petróleo Procesado (m3)	Mensual 1994-2013	MECON Info económica al día	--	15-30
natural	Industria	Producción de Gas Natural (m3)	Mensual 1994-2013	MECON Info económica al día	--	15-30
vacunos	Agricultura / Industria	Faena (En miles de cabezas)	Mensual 1990-2013	MECON Info económica al día	--	90-120
		Faena (En miles de cabezas)	Mensual 1990-2013	MinAgri Indicadores Bovinos mensuales	--	50-65
ovinos	Agricultura / Industria	Faena (En miles de cabezas)	Mensual 1994-2013	MECON Info económica al día	--	90-120
		Faena (En miles de cabezas)	Mensual 1990-2013	MinAgri Indicadores Ovinos mensuales, actuales	--	50-65
porcinos	Agricultura / Industria	Faena (En miles de cabezas)	Mensual 1994-2013	MECON Info económica al día	--	90-120
aves	Agricultura / Industria	Faena (En miles de cabezas)	Mensual 1994-2013	MECON Info económica al día	--	90-120
		Faena (En miles de cabezas)	Mensual 1994-2013	MinAgri Estadísticas de Aver. Info de Oferta y Dda Aviar	--	40-50
pescados	Pesca	Captura Pesquera (Tn)	Mensual 1992-2013	MECON Info económica al día	--	30-60
moluscos	Pesca	Captura Pesquera (Tn)	Mensual 1992-2013	MECON Info económica al día	--	30-60
crustaceos	Pesca	Captura Pesquera (Tn)	Mensual 1992-2013	MECON Info económica al día	--	30-60
cemento	Construcción / Industria	Producción (Miles de tn)	Mensual 1994-2013	MECON Info económica al día	--	30-60
hierro	Construcción / Industria	Producción (Miles de tn)	Mensual 1994-2013	MECON Info económica al día	--	30-60
acero	Construcción / Industria	Producción (Miles de tn)	Mensual 1994-2013	MECON Info económica al día	--	30-60
eléctrica	SSPP	Generación Energía Eléctrica (en GWh)	Mensual 1993-2013	MECON Info económica al día	--	15-30
gas_consumo	SSPP	Consumo de Gas Natural (millones de m3)	Mensual 1993-2013	MECON Info económica al día	--	15-30
agua	SSPP	Agua entregada por Aguas Argentinas (miles de m3)	Mensual 1993-2013	MECON Info económica al día	--	15-30
superficie	Construcción	Superficie Permisible (en miles de m2)	Mensual 1991-2013	INDEC - ISAC	--	20-30
publico	Adm. Pública	Gasto en Remuneraciones (EAIF)	Mensual 1993-2013	MECON	Defactadas (IPC City)	50-60
depósitos	Financiero	Depósitos + Préstamos Totales en el Sistema Financiero (En Pesos y en dólares)	Mensual 1996-2013	BCRA	Defactadas (IPC City)	15
merval	Financiero	Índice MerVal (índice promedio mensual)	Mensual 1991-2013	Yahoo finance	Defactadas (IPC City)	1
ventas	Comercio	Ventas de automóviles (nacional +importado)	Mensual 1995-2013	MECON	En unidades	5-10
		Ventas de automóviles (nacional)	Mensual 1994-2013	MECON	Empalme con serie nac + impo para el año 1994	5-10
transporte	Transporte y comunicaciones	Pasajeros serv. Ferroviario urbano	Mensual 1994-2013	MECON	Sumadas para un único índice en miles de pasajeros	30-45
		Pasajeros serv. Ferroviario interurbano	Mensual 1994-2013	MECON	Sumadas para un único índice en miles de pasajeros	30-45
		Pasajeros subterráneos Metrovias	Mensual 1994-2013	MECON	Sumadas para un único índice en miles de pasajeros	30-45
		Pasajeros serv. Transp. Automotor metropolitano	Mensual 1994-2013	MECON	Sumadas para un único índice en miles de pasajeros	30-45
		Pasajeros cabotaje aeronavegación comercial	Mensual 1994-2013	MECON	Sumadas para un único índice en miles de pasajeros	30-45
exportaciones	Comercio Exterior	Importaciones (En millones de US\$)	Mensual 1990-2013	INDEC	Defactadas (Precios Comercio Exterior)	20-25
		Índice de Precios de Comercio Exterior (Base 2004=100)	Trimestral 1990-2013	INDEC	Mensualizadas con Eviews	20-25 terminado el trim
importaciones	Comercio Exterior	Exportaciones (En millones de US\$)	Mensual 1990-2013	INDEC	Defactadas (Precios Comercio Exterior)	20-25
		Índice de Precios de Comercio Exterior (Base 2004=100)	Trimestral 1990-2013	INDEC	Mensualizadas con Eviews	20-25 terminado el trim
tdi	Impulsos	Índice de Precios de Comercio Exterior (Base 2004=100) - Exportaciones	Trimestral 1990-2013	INDEC	Mensualizadas con Eviews	20-25
		Índice de Precios de Comercio Exterior (Base 2004=100) - Importaciones	Trimestral 1990-2013	INDEC	Mensualizadas con Eviews	20-25 terminado el trim
capitales	Impulsos	Saldo de la Cuenta Capital y Financiera de Balanza de Pagos	Trimestral 1994-2013	INDEC	Mensualizadas con Eviews	60

2.2 Transformación de los datos

La estimación de un modelo factorial dinámico contempla la posibilidad de que haya persistencia en la relación entre las variables de las que se extrae el factor dinámico y supone que la relación entre ellas no es sólo contemporánea. Sin embargo, al igual que todos los modelos de series de tiempo, es necesario que se cumplan algunas condiciones específicas.

Se analizó la presencia de raíces unitarias mediante el test de Dickey-Fuller para cada una de las variables, diferenciando aquellas que fuera necesario para asegurarse la estacionariedad (ver anexo 1). A continuación se procedió a estandarizar las variables para neutralizar las diferencias en las unidades de medida de cada una de ellas, ya que de no hacerlo sus diferencias podrían influir en las cargas de cada variable dentro del factor común. La estandarización se realizó de acuerdo a la siguiente fórmula:

$$x_t = \frac{X_t - \bar{X}_t}{\widehat{sd}(X_t)}$$

Donde x_t es la variable estandarizada, cuya media es 0 y su desvío estándar es 1; \bar{X}_t es el promedio muestral entre 1994 y 2013 de la variable; \widehat{sd} es el desvío estándar muestral entre 1994 y 2013.

3. Estimaciones

3.1 Resultados Empíricos

La estimación del modelo factorial se hizo en dos etapas:

Etapla 1. Relaciones contemporáneas: Siguiendo la metodología de componentes principales, se estimaron las cargas correspondientes a la ecuación (1), contenidas en la matriz $\beta(L)$.

La tabla 2 presenta el vector de cargas sectoriales de cada indicador en el primer factor latente estimado. Se ve que en líneas generales todas las cargas son positivas. La razón por la cual algunas de ellas tienen cargas negativas puede deberse a que algunos indicadores (como por ejemplo tdi y capitales) corrigen el efecto bruto de otra variable.

TABLA 2. 1° ETAPA: CARGAS SECTORIALES ESTÁTICAS

Indicador	Carga Factorial
autos	0.01069
molienda	0.18507
energia	0.00728
faena	0.03054
pesca	-0.02912
construccion	0.16607
electronica	0.05361
gas_consumo	0.01097
agua	0.19046
superficie	0.03548
publico	0.14368
dp	-0.02524
merval	-0.00762
ventas	0.01915
exportaciones	0.34157
importaciones	0.01607
transporte	0.00902
tdi	0.01568
capitales	-0.01263

Etapa 2. Estimación dinámica: a partir del factor estático estimado en la etapa anterior, se estimó por máxima verosimilitud un modelo autorregresivo siguiendo la metodología de Box-Jenkins. El orden del modelo dinámico se definió siguiendo criterios de información y las funciones de autocorrelación del factor. El orden sugerido fue 2 para el componente autorregresivo y 1 para el de medias móviles; pero se incorporó también un componente autorregresivo de orden 12 acorde a la frecuencia mensual del factor, además de 12 variables dicotómicas para capturar los factores estacionales (Tabla 3). En el anexo 2 se presenta todo el modelo estimado en detalle con la evaluación de los residuos, cuya distribución ruido blanco aseguran la consistencia de los estimadores, junto con el cálculo de las raíces del modelo, que caen dentro del círculo unitario (lo que asegura la estabilidad del proceso autorregresivo estimado).

TABLA 3. 2° ETAPA: MODELO AUTORREGRESIVO

L.ar	L2.ar	L3.ar	L4.ar	L5.ar	L6.ar	L7.ar	L8.ar	L9.ar	L10.ar	L11.ar	L12.ar	L.ma
1.286***	-0.382	0.124	-0.153	0.126	-0.00595	-0.013	-0.0283	0.167	-0.174	0.206*	-0.156**	-0.620*
(0.365)	(0.292)	(0.125)	(0.113)	(0.11)	(0.111)	(0.117)	(0.104)	(0.111)	(0.135)	(0.121)	(0.0752)	(0.35)

D.MONTH1	D.MONTH2	D.MONTH3	D.MONTH4	D.MONTH5	D.MONTH6	D.MONTH7	D.MONTH8	D.MONTH9	D.MONTH10	D.MONTH11	Constante
-0.469***	-0.810***	-0.095	-0.127	0.270***	-0.0158	0.121	-0.0281	-0.161*	-0.00473	-0.221***	0.2
(0.0518)	(0.0672)	(0.0753)	(0.0839)	(0.087)	(0.0927)	(0.0941)	(0.084)	(0.0839)	(0.0738)	(0.0564)	(1.018)

Desvío estándar entre parentésis

*** Coeficientes significativos al 1%

** Coeficientes significativos al 5%

* Coeficientes significativos al 10%

La inclusión de 12 lags en la especificación del modelo responde a la naturaleza mensual del factor estimado. De hecho, sólo los lags 1, 11 y 12 resultaron significativos al 10 por ciento. Los coeficientes que capturan estacionalidad (D.month1 – D.month11 y constante) permiten corregir algunas oscilaciones periódicas del factor.

3.2 Identificación de Ciclos Económicos

El modelo se estimó utilizando series estandarizadas, por lo cual es preciso re-escalar⁹ el factor de forma tal que sea comparable con los niveles de otras series oficiales. Partiendo de la fórmula de estandarización de una variable,

$$\hat{f}_t = \frac{\hat{y}_t - \alpha}{\beta}$$

Donde α es la media de la variable; β es el desvío estándar; \hat{f}_t es el factor estandarizado estimado con la metodología de análisis factorial dinámico; y \hat{y}_t es la variable factorial original.

Despejando de la ecuación anterior, se puede estimar \hat{y}_t con la siguiente fórmula $\hat{y}_t = \alpha + \beta \hat{f}_t$ donde \hat{y}_t es la serie reescalada; \hat{f}_t es el factor estandarizado. Como el objetivo del factor es predecir el comportamiento de la serie oficial del EMAE, se tomaron los valores α (media) y β (desvío estándar) de la serie oficial del EMAE entre 1996 y 2013, bajo el supuesto de que la media y el desvío estándar observados en la serie oficial definen la realidad económica en ese período.

⁹ La utilidad de re-escalar la variable es evidente cuando se quiere proceder a una evaluación formal de la capacidad de pronóstico de cierta variable de interés (en este caso, el EMAE). Un camino alternativo hubiera sido estandarizar la serie del EMAE oficial y construir modelos con variables estandarizadas. Aunque la interpretación de una variable estandarizada no es compleja (su lectura se mide en cantidad de desvíos respecto de la media), el camino seguido en este trabajo permite una lectura de los datos más intuitiva.

De esta manera, se le asignan al factor los mismos parámetros de distribución que el EMAE (i.e. el mismo proceso generador de datos), y la esencia de la técnica se reduce a la capacidad que tiene de generar la misma muestra observada entre los años 1996 y 2013.

En el próximo capítulo se evalúa la técnica implementada desde dos enfoques: primero, para un subperíodo (1996 – 2007) se desarrolla una evaluación formal del método factorial dinámico, mediante la construcción de un modelo de pronóstico, a partir de la capacidad de predicción que tiene el factor sobre el comportamiento del EMAE, como una forma de validar la técnica utilizada; en segundo lugar, se realiza un análisis informal a partir del año 2008 comparando la evolución del factor y del EMAE.

4. Evaluación de Pronóstico

4.1 Período 1996-2007: Evaluación Formal

Esta evaluación formal permite validar la técnica de estimación utilizada. Se la instrumentó haciendo un modelo de predicción del EMAE a partir del factor estimado previamente. Para ello, se utilizó sólo el período 1996 y 2007, el que a su vez fue subdividido en dos submuestras:

- *Enero 1995 y Diciembre 2006*: período de estimación. Se utilizaron estos datos para estimar una regresión por máxima verosimilitud que explique la variación interanual del EMAE en función de “p” rezagos propios y “j” rezagos del factor latente.
- *Enero 2007 – Diciembre 2007*: período fuera de la muestra (*out-of-sample*). Este período no intervino en la estimación de los parámetros de la regresión, y se utilizó para medir la capacidad predictiva del modelo, a través del Error Cuadrático Medio.¹⁰

Se llevaron a cabo dos modelos de pronóstico para la primera submuestra (para ver los resultados de las estimaciones ver anexo 4):

En primer lugar, se estimó una regresión por mínimos cuadrados ajustados por heteroscedasticidad para proyectar las diferencias interanuales del logaritmo del EMAE en función del factor latente estimado, controlando por rezagos del EMAE para obtener residuos distribuidos

¹⁰ El período enero 2008 – diciembre 2013 fue excluido de la evaluación por la creciente disparidad entre los resultados obtenidos en este trabajo y los oficiales del EMAE. El sesgo no presenta un comportamiento sistemático que permita su estimación para la posterior corrección de los pronósticos, y se lo asigno a la falta de credibilidad de las cifras oficiales. Esta decisión está fundamentada, a su vez, en la corrección de las cifras anuales de crecimiento que fue posteriormente corregida cuando se publicaron las nuevas cifras de Cuentas Nacionales base 2004 (ver gráfico 2). Como hasta el momento no se cuenta con una serie de EMAE actualizada que permita calibrar la estimación del factor, se decidió eliminar estos años del pronóstico.

como ruido blanco y asegurar la consistencia de los indicadores. En términos generales, el modelo es de la siguiente forma:

$$m_t = a + \sum_{j=1}^6 \beta_j m_{t-j} + \sum_{h=0}^1 \delta_h f_{t-h} + \sum_{d=1}^{11} \theta_d D_{d,t} + e_t \quad ; \text{con } e_t \sim WN(0, \sigma_e^2)$$

Donde m es la diferencia 12-ava del logaritmo del EMAE oficial; m_{t-j} es el rezago “j” de la variable dependiente; f_{t-h} es el rezago “h” del factor latente estimado; $D_{d,t}$ es una variable dicotómica estacional (se incluyeron 11 variable categóricas para evitar la multicolinealidad, ya que también se estimó un término independiente, a); e_t es el residuo de la regresión, que se distribuye como ruido blanco, con media 0 y varianza σ_e^2

En segundo lugar, se estimó un modelo autorregresivo de orden 8 de las variaciones interanuales del EMAE. En términos generales,

$$m_t = a + \sum_{j=1}^8 \beta_j m_{t-j} + v_t \quad ; \text{con } v_t \sim WN(0, \sigma_v^2)$$

Donde m es la diferencia 12-ava del logaritmo del EMAE oficial; m_{t-j} es el rezago “j” de la variable dependiente; v_t es el residuo de la regresión, que se distribuye como ruido blanco, con media 0 y varianza σ_v^2 .

Utilizando los parámetros estimados, se procedió a la proyección de la variación interanual del EMAE bajo ambos modelos y las dos submuestras y se calculó, en cada caso, el error cuadrático medio (ECM), que mide el promedio del cuadrado de los errores de pronóstico - $(m_t - \hat{m}_t)^2$ -. El ECM penaliza simétricamente los desvíos por encima y por debajo del verdadero valor observado – aunque es sensible a la presencia de outliers¹¹:

$$ECM = \frac{1}{T} \sum_{t=1}^T (m_t - \hat{m}_t)^2$$

TABLA 4. EVALUACIÓN DE PRONÓSTICOS. FACTOR VS MODELO AUTORREGRESIVO.

	PERÍODO	
	2007	
	Obs.	ECM
Factorial	12	24.4155
Autorregresivo	12	24.4178

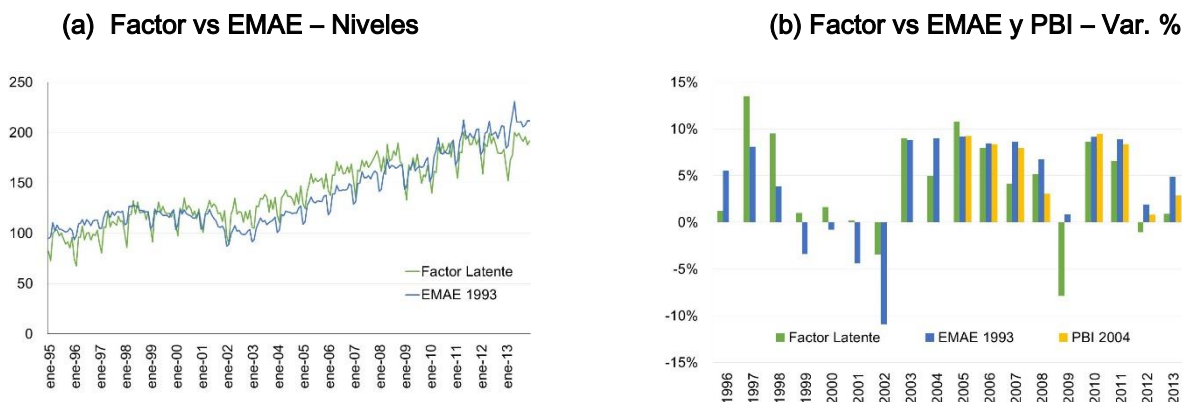
¹¹ Ahumada (2011).

La tabla 4 presenta la estimación del ECM para un horizonte de 12 meses (período *out-of-sample*). Allí puede verse que ambos modelos tienen un poder de predicción similar en términos de ECM. Sin embargo, el ajuste se inclina levemente a favor del modelo que incluye el valor presente y pasado del factor latente, ya que tiene un error cuadrático medio levemente inferior al autorregresivo estimado, brindando evidencia empírica que apoya los resultados obtenidos por Stock y Watson (2002). En su trabajo de investigación encontraron evidencia para Estados Unidos de que las proyecciones utilizando factores dinámicos superan – también en términos de ECM – a aquellos modelos que se basan en componentes autorregresivos o indicadores líderes.

4.2 Período 1996-2013: Evaluación Informal

El gráfico 2 muestra la evolución del factor estimado reescalado comparado con la evolución mensual del EMAE (panel -a-) y con las variaciones anuales del EMAE y del PBI base 2004¹² (panel -b-).

GRÁFICO 2: Factor Latente vs Series Oficiales



Allí puede verse la elevada correspondencia entre el factor latente y las series oficiales de PBI hasta el año 2007, momento a partir del cual el factor latente muestra, persistentemente un sesgo hacia abajo en comparación al EMAE. En particular, el factor muestra, a diferencia del EMAE y del PBI, una caída de la actividad económica en los años 2009 y 2012, y una tasa de crecimiento casi nula en el año 2013.

¹² Las Nuevas Cuentas Nacionales base 2004, hasta el momento sólo cuentan con una frecuencia trimestral y anual, y aún no se ha dado a conocer una serie histórica del Estimador Mensual. Adicionalmente el período relevado abarca el año 2004 hasta la actualidad.

La discrepancia entre las series oficiales y el factor estimado a partir del año 2008 podría reflejar un cambio estructural de la economía a partir de ese año o la distorsión de las cifras oficiales de crecimiento.

La primera opción (cambio estructural) podría tener como fuente de distorsión algunas de las siguientes (y que podrían, eventualmente, inducir a posibles líneas de investigación futuras):

- Modificación del valor poblacional del ponderador de alguno de los indicadores sectoriales utilizados para la construcción del factor dinámico;
- Aparición de nuevas actividades o nuevos eventos que ninguno de los indicadores utilizados incorporan;

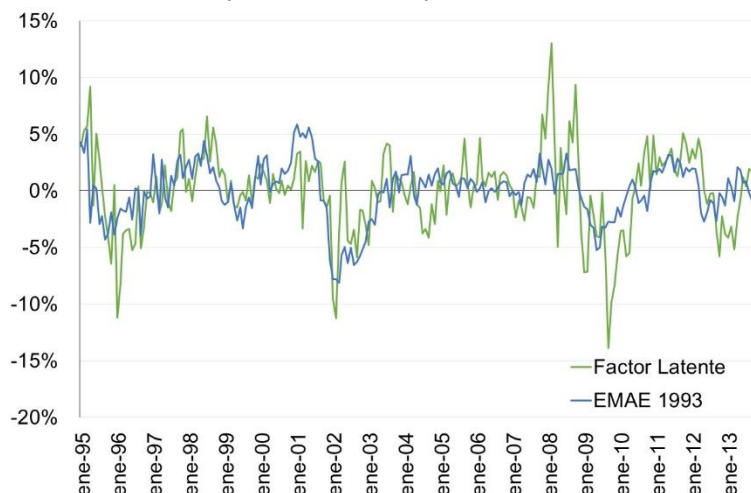
Aun teniendo en cuenta este sesgo sistemático hacia abajo en la estimación del factor, ambas series mensuales (EMAE y el factor dinámico) parecen identificar los mismos ciclos económicos, haciendo que la comparación entre ambas series se concentre en la estimación de la tendencia económica. Parecería que la tendencia del factor dinámico crece a una tasa anual más baja que la del EMAE.

La construcción de una serie de ciclos económicos para el factor latente y la serie del EMAE oficial se realizó aplicando el filtro de Hodrick-Prescott¹³ al logaritmo de cada serie y se tabularon los procesos recesivos y expansivos que se desprenden de cada una. Aquí un proceso recesivo (expansivo) fue definido como aquel en el que el nivel observado del indicador económico es inferior (superior) a su nivel tendencial. El gráfico 3 muestra las series de ciclos construidos para ambas series, y en la tabla 5 se identifican los períodos recesivos detectados en cada serie.

¹³ Ver Anexo 3.

GRÁFICO 3 FILTRO HODRICK-PRESCOTT. Factor vs EMAE

En desvíos porcentuales respecto de la tendencia.

**TABLA 5. PERÍODOS RECESIVOS. FACTOR VS EMAE.**

Años de Recesión			
	Factor	EMAE	Descripción
1	1996	1996	Tequila
2	1999	1999	Devaluaciones Emergentes
3	2002	2002-2003	Crisis Convertibilidad
4	2004	2004	Crisis Convertibilidad
5	2009-2010	2009-2010	Crisis Internacional y Campo
6	2012-2013	2012-2013	Control de Cambios

Ambas series identifican claros períodos de expansión desde 1996 a 1999, año a partir del cual comienza a avizorarse la crisis de la convertibilidad, que tardó hasta el año 2005 hasta recuperar los niveles tendenciales de producción económica. El crecimiento entre el año 2005 y 2008 protagonizó un nuevo período expansivo en ambas series, seguido de la crisis internacional (y del campo internamente) entre 2009 y 2010; y luego un último período recesivo entre 2012 y 2013 asociado a la imposición del control de cambios.

5. Conclusiones

En este trabajo se aplicó una técnica de estimación de un modelo factorial dinámico para procesar e interpretar la información contenida en un conjunto de series económicas sectoriales. El modelo utilizado permite sintetizar toda la información en un único indicador global que pondera cada uno de los sectoriales de acuerdo a su carga individual en el promedio de la economía, minimizando la pérdida de información que significaría eliminar variables explicativas por no contar con una cantidad suficiente de observaciones temporales.

La característica distintiva de este modelo es la incorporación de la persistencia temporal de los indicadores sectoriales en el factor latente, levantando el supuesto de que las relaciones entre las variables son sólo contemporáneas.

Enmarcado en un contexto de desconfianza en las estadísticas oficiales, el objetivo fue estimar un nuevo indicador económico, utilizando un enfoque estadístico, que funcione como parámetro para evaluar la calidad y, - dado que para su construcción se utilizaron otros indicadores elaborados por organismos oficiales de estadísticas -, la consistencia interna de las estadísticas oficiales. Se encontró evidencia empírica que apoya la presentada por Coremberg (2014), de que el crecimiento económico habría sido sobre-estimado por los organismos oficiales de estadísticas a partir del año 2008.

La técnica utilizada aquí es universal y permitiría conseguir el doble objetivo de adelantarse a la publicación del EMAE, siempre que se la complemente con técnicas estadísticas adicionales que permitan trabajar con paneles desbalanceados, por lo que la utilidad de este indicador no es exclusiva al contexto coyuntural actual.

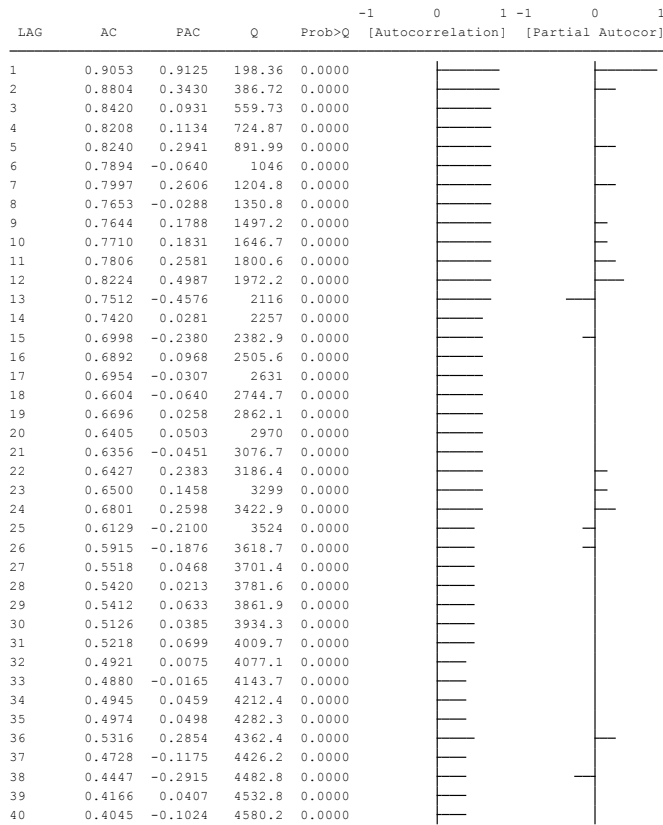
Por último, luego del cambio de Cuentas Nacionales a comienzos del año 2014, se perdió la evolución histórica del EMAE entre 1994 y 2014. La utilización de este modelo no sólo permite contar con información en tiempo real, sino también podría contribuir a la reconstrucción de la historia del EMAE, si se re-escala más frecuentemente (e.g. trimestralmente) utilizando los datos del PBI.

ANEXO 1. REDUCCIÓN DE LA DIMENSIÓN DE VARIABLES Y ESTACIONARIEDAD

Nombre	Variable agrupadora	Veces de diferenciación
autos	autos	1
soja	molienda	0
trigo	molienda	0
maíz	molienda	0
girasol	molienda	0
cebada	molienda	0
crudo	energía	0
procesado	energía	0
natural	energía	0
vacunos	faena	1
ovinos	faena	1
porcinos	faena	1
aves	faena	1
pescados	pesca	0
moluscos	pesca	0
crustaceos	pesca	0
cemento	construccion	0
hierro	construccion	0
acero	construccion	0
electrica	electrica	1
gas_consumo	gas_consumo	1
agua	agua	0
superficie	superficie	0
publico	publico	0
depósitos	dp	1
merval	merval	1
ventas	ventas	1
transporte	transporte	0
exportaciones	expo	0
importaciones	impo	1
tdi	tdi	1
capitales	capitales	0

ANEXO 2. MODELO FACTORIAL DINÁMICO: ETAPA 2. METODOLOGÍA BOX-JENKINS.

1.1 Correlograma del Factor



1.2 Evaluación de las raíces del Proceso Autorregresivo estimado para el Factor.

Characteristic roots of AR-polynomial

Characteristic roots	Modulus	Period
.9964887	.996489	
.7364155 + .4930527i	.886233	10.6
.7364155 - .4930527i	.886233	10.6
-.4983179 + .7024575i	.861259	2.87
-.4983179 - .7024575i	.861259	2.87
-.8100694 + .2872132i	.859479	2.24
-.8100694 - .2872132i	.859479	2.24
.8539183	.853918	
.3565611 + .7522503i	.832476	5.57
.3565611 - .7522503i	.832476	5.57
-.06686348 + .7806669i	.783525	3.79
-.06686348 - .7806669i	.783525	3.79

Characteristic roots of MA-polynomial

Characteristic roots	Modulus	Period
.6200775	.620077	

Todas las raíces caen dentro del círculo unitario, lo que asegura la estacionariedad del proceso ARMA estimado

1.3 Análisis de los residuos del modelo dinámico. Test Ljung – Box.

El análisis de los residuos muestra que no es posible rechazar la hipótesis nula de que los residuos poseen una distribución ruido blanco. La ausencia de autocorrelación de los residuos garantiza la consistencia de los estimadores.

LAG	AC	PAC	Q	Prob>Q	-1	0	1	-1	0	1
						[Autocorrelation]	[Partial Autocor]			
1	-0.0722	-0.0725	1.2629	0.2611						
2	0.0410	0.0393	1.6707	0.4337						
3	-0.0176	-0.0189	1.7463	0.6267						
4	0.0390	0.0366	2.1185	0.7140						
5	-0.0712	-0.0687	3.3653	0.6439						
6	-0.0086	-0.0152	3.3836	0.7594						
7	0.0070	0.0091	3.3956	0.8462						
8	-0.0043	-0.0112	3.4002	0.9068						
9	-0.0613	-0.0596	4.3411	0.8876						
10	-0.0199	-0.0295	4.4407	0.9253						
11	-0.0442	-0.0479	4.9348	0.9343						
12	0.0042	-0.0032	4.9392	0.9600						
13	-0.1483	-0.1543	10.547	0.6487						
14	0.2160	0.2114	22.485	0.0692						
15	-0.1224	-0.1280	26.338	0.0346						
16	0.0433	0.0320	26.822	0.0435						
17	-0.0008	0.0144	26.822	0.0607						
18	0.0051	-0.0367	26.829	0.0823						
19	-0.0287	0.0017	27.045	0.1036						
20	0.0322	0.0041	27.316	0.1266						
21	-0.0698	-0.0831	28.604	0.1238						
22	-0.0069	-0.0192	28.616	0.1561						
23	0.1391	0.1663	33.779	0.0684						
24	-0.0393	-0.0803	34.193	0.0813						
25	-0.0526	-0.0343	34.938	0.0893						
26	0.0048	-0.0359	34.944	0.1129						
27	-0.1325	-0.1257	39.714	0.0545						
28	0.0374	-0.0373	40.096	0.0648						
29	-0.0904	-0.0301	42.338	0.0523						
30	0.0265	-0.0488	42.531	0.0644						
31	0.0435	0.0736	43.056	0.0734						
32	0.0291	-0.0019	43.292	0.0878						
33	0.0071	0.0324	43.306	0.1081						
34	0.0095	-0.0182	43.331	0.1311						
35	0.0517	0.0913	44.086	0.1396						
36	-0.0107	0.0118	44.118	0.1659						
37	0.1447	0.0976	50.087	0.0739						
38	-0.1290	-0.1319	54.855	0.0377						
39	-0.0010	-0.0284	54.855	0.0474						
40	0.0008	-0.0566	54.856	0.0590						

ANEXO 3. FILTRO HODRICK-PRESCOTT.

Se parte de la descomposición de una serie en los componentes: estacionalidad (S_t), tendencia (T_t), ciclo (C_t) y error.

$$Y_t = S_t T_t C_t E_t$$

Tomando logaritmos:

$$y_t = s_t + \tau_t + c_t + e_t$$

Donde $i_t = \ln(I_t)$ para $i = Y, S, T, C, E$.

La estimación de la estacionalidad puede realizarse mediante algún mecanismo estandarizado (ARIMA X11 o ARIMA X12) para obtener $y^{sa}_t = y_t - s_t$.

Por lo tanto: $y^{sa}_t = \tau_t + c_t + e_t$

El filtro Hodrick-Prescott permite estimar el componente tendencial, τ_t , y a partir de allí, el cíclico.

Dadao un valor positivo λ , se calcula el componente estacional resolviendo la siguiente minimización:

$$\min \sum_{t=1}^T (y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} [(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2$$

Donde siempre: $\sum_{t=1}^T (y_t - \tau_t) = 0$ que muestra que la tendencia calculada pasa por el “centro” de la serie.

Luego de la estimación de la tendencia, se puede estimar el componente cíclico (que incluye los errores), $y^{sa}_t - \tau_t = c_t$

ANEXO 4. MODELOS DE PRONÓSTICO

En este anexo se presentan los valores estimados para los parámetros de ambos modelos de pronóstico utilizados en el capítulo 4.

Variables	Factor	Autorregresivo
	varemae	varemae
EMAE (t-1)	0.877*** (0.119)	0.941*** (0.0827)
EMAE (t-2)	-0.0405 (0.128)	-0.0765 (0.136)
EMAE (t-3)	0.553*** (0.112)	0.435*** (0.135)
EMAE (t-4)	-0.512*** (0.1)	-0.403*** (0.132)
EMAE (t-5)	0.347*** (0.127)	0.310** (0.126)
EMAE (t-6)	-0.276*** (0.0927)	-0.172 (0.106)
EMAE (t-7)		-0.264* (0.139)
EMAE (t-8)		0.171 (0.114)
Factor (t)	0.0991* (0.0561)	
Factor (t-1)	-0.120** (0.0482)	
D1	-0.00277 (0.0108)	
D2	-0.000503 (0.00945)	
D3	0.0011 (0.011)	
D4	0.0000679 (0.0125)	
D5	-0.000675 (0.0104)	
D6	0.00234 (0.00989)	
D7	-0.000782 (0.01)	
D8	0.00201 (0.0107)	
D9	0.000287 (0.0115)	
D10	-0.000338 (0.00917)	
D11	-0.00097 (0.0094)	
Constant	0.00226 (0.00902)	0.0327 (0.0301)
Observations	126	126

6. Referencias

Ahumada, H. (2011). *Break Detectability and Mean Square Forecast Error Ratios for Selecting Estimation Windows*. Journal of Forecasting.

CFNAI (2013) *Background on the Chicago Fed National Activity Index*. The Chicago Fed.

Coremberg, A. (2014). *Measuring Argentina's GDP Growth. Myth and facts*. World Economics, 15-1.

Cuevas, A. y Quilis, E. (2012). *A factor analysis for the Spanish economy*. Journal of The Spanish Economic Association.

Elosegui, P., Garegnani, L., Lanteri, y Blanco E. (2008). *Indicadores agregados de actividad económica para Argentina: el método de componentes principales*. BCRA, Ensayos Económicos. 51

Geweke, J. (1977). *The Dynamic Factor Analysis of Economic Time Series*. Latent Variables in Socio-Economic Models, ed. by D.J. Aigner and A.S. Goldberger, Amsterdam: North-Holland.

Hodrick, R. y Prescott, E. (1997). *Post-war U.S. business cycles: An empirical investigation*. Journal of Money, Credit and Banking, 29(1), 1-16.

Mariano R, Murasawa Y (2003). *A new coincident index of business cycles based on monthly and quarterly series*. J Applied Economics.

Sargent, T.J., y C.A. Sims (1977). *Business Cycle Modeling Without Pretending to Have Too Much A-Priori Economic Theory*. New Methods in Business Cycle Research, ed. by C. Sims et al., Minneapolis: Federal Reserve Bank of Minneapolis.

Stock, J. y Watson, M. (2010). *Dynamic Factor Models*. National Bureau of Economic Research.

Stock, J. y Watson, M. (2002). *Macroeconomic Forecasting Using Diffusion Indexes*. Journal of Business & Economic Statistics.

UTDT (2015). *Modificación metodológica del Índice Líder*. Centro de Investigación en Finanzas, Universidad Torcuato Di Tella.