



División de Política Monetaria
Gerencia de Análisis Macroeconómico
Departamento de Análisis de Coyuntura

Uso de modelos econométricos para la proyección de inflación en el Banco Central de Chile

Guillermo Carlomagno y Johanna Torres

Febrero de 2020

I. Objetivo

El objetivo de esta minuta es describir los procedimientos utilizados por el Banco Central para realizar proyecciones de inflación a corto plazo (uno a seis meses), y predicciones de precios de alimentos a mediano plazo (seis meses a un año). El caso de los precios de alimentos es especial debido a la dificultad de los modelos estructurales de replicar las dinámicas irregulares que presentan los alimentos y que pueden predecirse mejor con modelos empíricos. En esta minuta no se incluyen resultados de predicciones ni comparaciones empíricas entre procedimientos alternativos.

En la sección II se describen los modelos de corto plazo y en la III los de mediano plazo para los alimentos.

II. Modelos de corto plazo

El conjunto de modelos utilizados con ese fin puede dividirse en dos grupos: procedimientos directos con los que se proyectan series agregadas de interés (IPC, IPC SAE e IPC sin volátiles) y procedimientos indirectos, con los que se proyectan subclases del IPC y agregados intermedios, cuyas proyecciones se agregan para obtener la de los agregados. En adelante, llamaremos “componentes” tanto a las sub-clases como a los agregados intermedios, que representan un nivel de agregación menor que los agregados de interés.

Los procedimientos indirectos son de interés para el Banco por dos razones. En primer lugar, incluso si el único objetivo fuese predecir el agregado, el clásico debate entre métodos directos e indirectos (ver Kohn, 1982; Lütkepohl, 1984; Rose, 1977; Tiao y Guttman, 1980; y Wei y Abraham, 1981, entre otros) indica que, en teoría, los procedimientos directos son superiores sólo cuando los datos satisfacen condiciones muy específicas (ver Kohn, 1982 y Lütkepohl, 1987), que en general no se cumplen en la práctica. Aplicaciones recientes de procedimientos indirectos cuando el único objetivo es predecir el agregado pueden verse en Bermingham y D'Agostino (2014), Espasa y Mayo (2013), Hendry y Hubrich (2011), entre otros.

En segundo lugar, independientemente de la capacidad para predecir el agregado, el análisis y las proyecciones de los componentes son de interés para la política monetaria. Tanto es así que, con la misma perspectiva para la inflación agregada, distintas trayectorias esperadas de los



componentes podrían derivar en decisiones de política monetaria opuestas. Por ejemplo, dado un nivel de inflación cercano a la meta, y un determinado nivel de actividad real, no es indiferente para el banco que los precios de todos los componentes estén creciendo a una tasa similar, a que los alimentos no procesados y la energía se ubiquen muy por encima de la meta y los servicios por debajo.

Es claro entonces que la predicción de los componentes es un insumo esencial para la política monetaria. No obstante, la desagregación presenta tres desafíos: i) En primer lugar, es necesario decidir acerca del nivel “óptimo” de desagregación, si es que tal nivel existe. ii) Segundo, como se ilustra en la sección 0, los componentes del IPC chileno son, en general, significativamente más “ruidosos” que el IPC agregado. Ello se debe a que se ven más afectadas que el IPC por errores de medida, shocks de oferta sectoriales, cambios metodológicos, etc. En los modelos econométricos, estos eventos se reflejan en la aparición de datos atípicos y quiebres de la media incondicional, por lo que es necesario contar con procedimientos de estimación y predicción robustos a esos eventos. iii) En tercer lugar, se presenta el desafío de identificar qué variables macroeconómicas son relevantes para predecir cada uno de los componentes, y si existe algún componente que contenga información adelantada para predecir otro.

En esta minuta el problema i se resuelve utilizando los niveles de desagregación que, operativamente, resultan de interés para el análisis detallado de la inflación. Como se indica en la sección **¡Error! No se encuentra el origen de la referencia.**, el Banco considera un total de 54 series, que incluyen subclases, agregados intermedios y los agregados de interés.

Como se describe en la sección 0, los problemas ii y iii se resuelven de manera integral mediante la utilización del algoritmo de selección automática *Autometrics* (ver Doornik, 2009).

II.1 Metodología

La estrategia empírica utilizada para los procedimientos indirectos consiste en estimar un modelo uni-ecuacional para cada uno de los componentes considerados y luego agregar las predicciones para obtener predicciones de los agregados de interés (IPC, IPC SAE, e IPC sin volátiles) desde uno a seis meses hacia adelante. Para los procedimientos directos, se estiman modelos uni-ecuacionales para los agregados de interés y se obtienen las predicciones directas.

Como se indicó en la Introducción, la construcción de los modelos se realiza con el algoritmo de selección de modelos *Autometrics*. Partiendo de un modelo general que incluye un conjunto amplio de potenciales regresores (*general unrestricted model*, GUM) y utilizando una metodología de árboles de decisión de senderos múltiples, *Autometrics* reduce el modelo inicial a una representación final válida¹, más simple que la inicial.

Dentro de este marco general de modelación, se consideran tres estrategias alternativas para cada una de las 54 series consideradas, que difieren en los potenciales regresores incluidos en la

¹ La validez de la representación final se asegura mediante contrastes de congruencia y *encompassing* (ver Bontemps y Mizon, 2008; Hendry et.al., 2008; y Doornik, 2009).



especificación inicial, GUM. Las estrategias son: a) incluir únicamente rezagos de la variable dependiente, b) incluir rezagos de la variable dependiente y de los restantes componentes, c) incluir rezagos de la variable dependiente, de los restantes componentes y de las 10 variables macroeconómicas descritas en la sección **¡Error! No se encuentra el origen de la referencia..**

La ecuación 1 representa el GUM del caso c, el caso a y b se obtienen a partir de restricciones cero en alguno de sus parámetros.

$$[1 - \Phi(L)]\Delta y_{jt} = c + \sum_{i=1}^{11} \gamma_i S_{it} + \sum_{n=1, n \neq j}^N \lambda_n(L) \Delta y_{n,t} + \sum_{m=1}^Q \theta_m(L) \Delta x_{m,t} + e_{j,t}, \quad (1)$$

donde Δy_{jt} la diferencia logarítmica de la serie j , para $j = 1, \dots, 54$. $\Phi(L)$, es un polinomio auto-regresivo de orden 13, con los coeficientes 6 a 11 iguales a cero, $\lambda_n(L)$ y $\theta_m(L)$ son polinomios auto-regresivos de orden 5, S_{it} son variables estacionales artificiales centradas, $\Delta x_{m,t}$ es la diferencia logarítmica de la variable macroeconómica m , para $m = 1, \dots, 10$, y $e_{j,t}$ es un ruido blanco. Las variables macroeconómicas incluyen variables locales y externas (ver sección **¡Error! No se encuentra el origen de la referencia..**). Los elementos relevantes de la ecuación se seleccionan mediante Autometrics.

Mientras que en la estrategia a los polinomios $\lambda_n(L)$ y $\theta_m(L)$ se restringen para que todos sus parámetros sean cero, en la estrategia b se restringe solamente el polinomio $\theta_m(L)$.

Cuando Δy_{jt} es alguno de los agregados de interés, la estrategia b coincide con la propuesta de Hendry y Hubrich (2005, 2011), y la c es una ampliación de dicha propuesta en la que se agregan variables macroeconómicas. Hendry y Hubrich demuestran que, al incluir los retardos de los componentes en la ecuación del agregado, desaparecen las limitaciones teóricas de los procedimientos de predicción directos mencionadas en la Introducción. Asimismo, los autores indican que, en muestras finitas, su estrategia puede ser superior a los procedimientos indirectos debido a la reducción de la incertidumbre de estimación.

Dado que dentro de las 54 series consideradas hay diversos agregados intermedios y subclases individuales, las predicciones indirectas de los agregados de interés pueden obtenerse en base agregaciones diversas. La práctica actual del Banco es obtener las predicciones puntuales mediante el método directo, usando la estrategia c (propuesta de Hendry y Hubrich ampliada con variables macroeconómicas) y utilizar las predicciones indirectas para construir intervalos de predicción de uso interno.

El hecho de que todas las variables del lado derecho de la ecuación estén rezagadas al menos un período asegura su exogeneidad y la validez del modelo como instrumento de predicción (ver Hendry, 2017).

Para predicciones a horizontes mayores a un mes, los valores de las variables exógenas son desconocidos. A fin de evitar la construcción de modelos específicos para cada variable exógena



se adopta el procedimiento de predicción directa multi-período (ver Chevillon y Hendry, 2005)². El mismo consiste en estimar un modelo distinto para cada horizonte h fijando el orden mínimo de los polinomios $\lambda_n(L)$ y $\theta_m(L)$ en h (y el máximo en $h - 4$).

Asimismo, debido al calendario de publicación de cifras oficiales, el primer rezago de algunas variables exógenas no está disponible. En esos casos, y solo para el primer rezago, se utilizan las proyecciones internas del Banco.

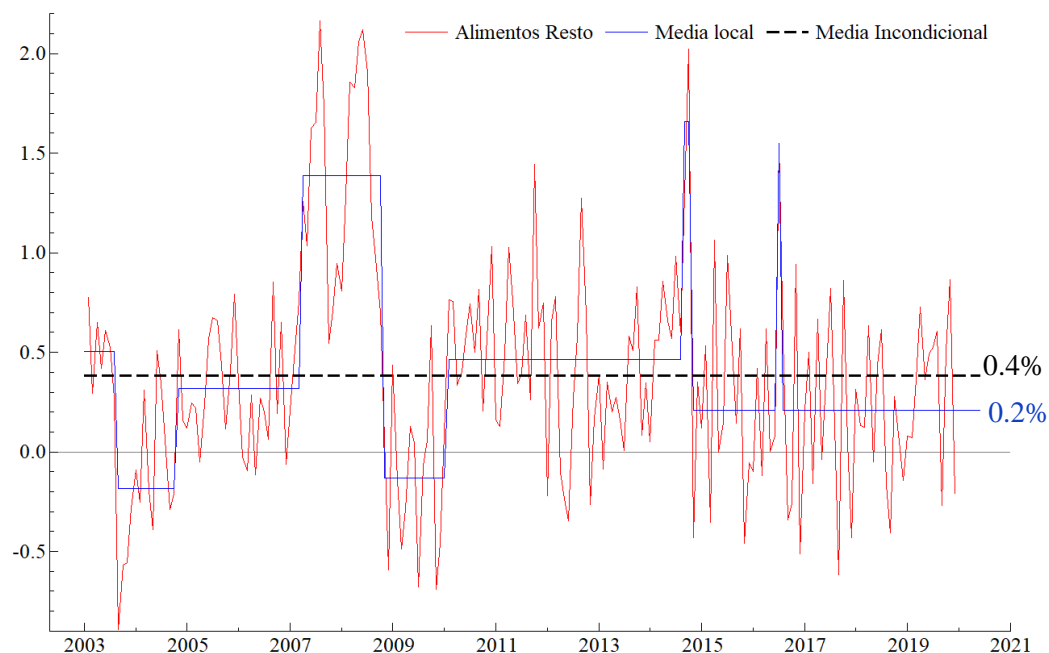
Datos atípicos y quiebres de nivel

Como la mayoría de los modelos econométricos dinámicos, la ecuación 1 contiene un mecanismo de ajuste al equilibrio, en el sentido que las predicciones que de ella se derivan (la esperanza condicional) convergen a la esperanza incondicional, cuya estimación coincide con la media muestral. Por tanto, si la media incondicional sufre quiebres esporádicos en el período de estimación y/o en el de predicción, es necesario tenerlos en cuenta para evitar sesgos sistemáticos en la predicción (ver Hendry, 2006 y Castle, et.al., 2015).

Como se indicó en la Introducción, las series desagregadas del IPC chileno están especialmente expuestas a la ocurrencia de valores atípicos y quiebres de nivel. El gráfico 1 ilustra este problema; la media local del agregado intermedio “alimentos resto” (línea azul) ha ido sufriendo quiebres a lo largo del tiempo y en la actualidad se ubica significativamente por debajo de la media incondicional (línea negra discontinua). Si no se tuviesen en cuenta dichos quiebres tendríamos dos problemas: primero la estimación de los parámetros del modelo sería sesgada, segundo, y más importante, las predicciones serían sesgadas pues convergerían a la línea negra discontinua.

Gráfico 1. Ejemplo de inestabilidad de la media local en los componentes del IPC. Variación mensual de “alimentos resto” (en porcentaje, período 2003.2-2019.12) y cambios de la media local estimados son IIS y SIS.

² Para una discusión detallada de métodos directos y recursivos de predicción, ver también Chevillon (2007). Ben et. al., (2012) proponen una tercera alternativa que combina ambos métodos.



Nota: El agregado intermedio “alimentos resto” se obtiene excluyendo frutas, verduras y bebidas no alcohólicas del sub-agregado “alimentos”.

Para abordar este problema, se adoptan las técnicas de saturación Impulse Indicator Saturation (IIS) y Step Indicator Saturation (SIS) propuestas y analizadas por Doornik et.al. (2013) y Castle et.al. (2015) y Johansen y Nielsen (2009).

El control de los datos atípicos y quiebres de nivel en la muestra reduce el error estándar del modelo, lo que llevaría a sub-estimar la incertidumbre asociada a la predicción. Una solución simple a este problema es agregar la varianza del residuo con la de las intervenciones (línea azul del gráfico 1) para el cálculo de la incertidumbre de la predicción.

El problema de la posible ocurrencia de quiebres de la media en el período de predicción se mitiga considerando la estrategia de robustificación propuesta por Hendry (2006) y Castle, et.al. (2015). Bajo esta estrategia, las predicciones se obtienen de un modelo sobre-diferenciado (con los parámetros estimados en el modelo correcto), lo que evita errores sistemáticos cuando ocurren quiebres. El costo de la sobre-diferenciación es que se incrementa la varianza del error de predicción cuando no ocurren quiebres.

Las tres posibilidades descritas para la ecuación 1 y las distintas alternativas de agregación de los componentes y los agregados intermedios entregan un conjunto de predicciones que son monitoreadas periódicamente.



II.2 Datos

La tabla 1 contiene el listado de todos los componentes y agregados de interés que se proyectan regularmente utilizando los métodos descrito arriba. La tabla 2 indica las variables macroeconómicas consideradas en los modelos (variables x de la ecuación 1).

Tabla 1. Listado de agregados, componentes y sub-agregados considerados

	Componentes	Ponderaciones. Base 2018	Notas
1	IPC	100.00	
2	IPCSAE	73.16	
3	Bienes SAE	27.30	
4	Servicios SAE	45.86	
5	Alimentos	19.30	
6	Energía	7.54	
7	Resto de alimentos	16.07	
8	Pan y cereales	4.03	
9	Carnes, pescados y mariscos	4.96	
10	Productos lácteos y huevos	2.42	
11	Bebidas	2.16	
12	Otros alimentos	2.48	
13	Frutas y verduras frescas	3.24	
14	Electricidad	2.27	
15	Combustibles	5.27	
16	Bebidas alcohólicas	2.87	
17	Cigarrillos	1.91	
18	Vestuario y calzado	4.06	
19	Vivienda	1.14	
20	Equipamiento de vivienda electrónico	1.63	
21	Resto de equipamiento de vivienda	4.37	
22	Bienes de salud	5.28	
23	Automóvil nuevo	2.86	
24	Otros bienes de transporte	0.80	
25	Paquete turístico	0.77	
26	Artículos de recreación	1.62	
27	Arriendo	5.53	
28	Agua potable	1.66	
29	Gasto común	1.37	
30	Otros servicios de vivienda	1.37	
31	Servicios domésticos	2.65	
32	Servicios Salud	5.27	
33	Aéreo	0.75	
34	Bus	0.49	
35	Otros servicios de transporte	3.07	
36	Servicios de telecomunicaciones	4.84	
37	Servicios de recreación	2.59	
38	Educación	6.60	
39	Servicios de alimentación	5.87	
40	Gasto financiero	0.23	
41	Otros servicios	3.57	
42	Servicios Administrados e Indexados	30.44	En base a Marcel, Medel y Mena (2017)
43	Servicios Resto	15.42	En base a Marcel, Medel y Mena (2017)
44	Elasticidad Salario Alta	32.19	En base a VAR de Pinto y Contreras (2016)
45	Servicios correlacionados con salarios	24.47	23 subclases más correlaciones con CMO
46	IPC s/volátiles	65.15	En base a Carlomagno y Sansone (2019)
47	Bienes s/volátiles	26.72	En base a Carlomagno y Sansone (2019)
48	Servicios s/volátiles	38.42	En base a Carlomagno y Sansone (2019)
49	IPC Volátiles	34.85	En base a Carlomagno y Sansone (2019)
50	Bienes s/volátiles ex Alimentos	17.53	En base a Carlomagno y Sansone (2019)
51	Alimentos no Volátiles	9.20	En base a Carlomagno y Sansone (2019)
52	Alimentos Volátiles	10.11	En base a Carlomagno y Sansone (2019)
53	Energía Volátiles	7.50	En base a Carlomagno y Sansone (2019)
54	Resto Volátiles	17.25	En base a Carlomagno y Sansone (2019)

Fuente: INE y BCChile



Tabla 2. Listado de agregados, componentes y sub-agregados considerados

Variable	Fuente
IMACEC no minero	Banco Central de Chile
Costo de mano de obra (ICMO)	Instituto Nacional de Estadísticas
Tasa de desempleo	Instituto Nacional de Estadísticas
Tipo de cambio nominal	Banco Central de Chile
WTI	Bloomberg
Brent	Bloomberg
Índice de precios externos (IPE)	Banco Central de Chile
Índice de precios de alimentos (FAO)	FAO
Fed Fund Rate (FFR)	Bloomberg
Producción industrial global	Bloomberg
TPM de Chile	Banco Central de Chile

III. Modelos de cointegración para proyecciones de mediano plazo de Alimentos

Dada la dificultad de los modelos estructurales para proveer proyecciones precisas para la inflación de alimentos, dichos modelos se informan con las proyecciones de modelos econométricos no sólo para el corto plazo, como en la mayoría de las variables, sino que también en horizontes más extensos. Dada esta necesidad de proyecciones a plazos más largos, la estrategia de proyección para las series de inflación de los distintos componentes de “alimentos resto” considera modelos que incorporan simultáneamente relaciones de corto y largo plazo. En particular, se utilizan modelos cointegración que permiten explicar dinámicas y relaciones de corto plazo entre variables.

En términos generales, los modelos utilizados para la proyección de los precios de alimentos pueden expresarse como:

$$\Delta y_t = \beta \Delta x_t^{cp} - \theta_1 (y_{t-1} - \theta_2 x_{t-1}^{lp}) + \epsilon_t, \quad (2)$$

donde y_t es la variable a proyectar; x_t^{cp} contiene las variables que explican el comportamiento de corto plazo de y_t (x_t^{cp} puede contener rezago de y_t); x_t^{lp} contiene las variables que definen las relaciones de largo plazo de y_t ; θ_1 representa la velocidad de ajuste al equilibrio de y_t , θ_2 el vector de cointegración, y ϵ_t corresponde a un ruido blanco. La estimación de los parámetros de largo plazo (θ_2) se realiza mediante el procedimiento de Engle y Granger.



La ecuación 2 se estima para los precios de carnes; lácteos y sus derivados; pan y cereales; azúcar y cacao; y aceites y grasas. La tabla 3 presenta las variables utilizadas en la ecuación de largo plazo (x_t^{lp}) de cada modelo. Para cada especificación, se incluye adicionalmente una variable *dummy* con el objeto de controlar el efecto de la crisis financiera durante el periodo 2008. Exceptuando la *dummy*, todas las variables se expresan en términos de logaritmos naturales. El número de rezagos de la ecuación de corto plazo se determina utilizando el criterio de información de Schwartz.

Se utilizan datos de frecuencia trimestral, desde el 2000 hasta el último trimestre disponible. Para los datos de precios, se utilizan los sucesivos empalmes de las canastas del IPC elaboradas por el Instituto Nacional de Estadísticas (INE) para los años 1998, 2008, 2009, 2013 y 2018. Para la agregación de las subclases, se utilizan los ponderadores relacionados a la última canasta disponible (2018).

Tabla 3. Variables utilizadas para la proyección de la inflación de cada grupo de alimentos³

Grupo	Variable
Carnes	IPC de la clase carnes
	Índice FAO carnes
	IPC del grupo granos y cereales
	IPC tarifas eléctricas
	CMO
Lácteos y derivados	Tipo de cambio observado
	IPC de la clase lácteos y derivados
	Índice FAO lácteos
	IPC del grupo granos y cereales
	CMO
Pan y Cereales	Tipo de cambio observado
	IPC de la clase granos y cereales
	Índice FAO cereales
	IPC tarifas eléctricas
	IPC combustibles
Azúcar y Cacao	CMO
	Tipo de cambio observado
	IPC de la clase azúcar y cacao
	Índice FAO azúcar
	Brent
Aceites y Grasas	CMO
	Tipo de cambio observado
	IPC de la clase aceites y grasas
	Índice FAO aceite
	IPC tarifas eléctricas
Aceites y Grasas	IPC combustibles
	CMO
	Tipo de cambio observado

³ Los contrastes de raíces unitarias ADF y Phillips-Perron indican que todas las variables son I(1).



Referencias

- Bermingham, Colin, and Antonello D'Agostino. "Understanding and Forecasting Aggregate and Disaggregate Price Dynamics." *Empirical Economics* 46, no. 2 (2014): 765-88.
- Ben Taieb, Souhaib, and Rob Hyndman. "Recursive and Direct Multi-Step Forecasting: The Best of Both Worlds." Monash University, Department of Econometrics and Business Statistics, 2012.
- Bontemps, Christophe, and Grayham E Mizon. "Encompassing: Concepts and Implementation." *Oxford Bulletin of Economics and Statistics* 70 (2008): 721-50.
- Castle, Jennifer L, Michael P Clements, and David F Hendry. "Robust Approaches to Forecasting." *International Journal of Forecasting* 31, no. 1 (2015): 99-112.
- Castle, Jennifer L, Jurgen A Doornik, David F Hendry, and Felix Pretis. "Detecting Location Shifts During Model Selection by Step-Indicator Saturation." *Econometrics* 3, no. 2 (2015): 240-64.
- Chevillon, Guillaume, and David F Hendry. "Non-Parametric Direct Multi-Step Estimation for Forecasting Economic Processes." *International Journal of Forecasting* 21, no. 2 (2005): 201-18.
- Chevillon, Guillaume. "Direct Multi-Step Estimation and Forecasting." *Journal of economic surveys* 21, no. 4 (2007): 746-85.
- Doornik, J.A. "Autometrics." Chap. 4 In *Castle and Shephard, the Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry.*, 88-121: Oxford University Press, 2009.
- Doornik, Jurgen A, David F Hendry, and Felix Pretis. "Step-Indicator Saturation." *University of Oxford. Working Paper*, no. 658 (2013).
- Espasa, A, and I Mayo. "Forecasting Aggregates and Disaggregates with Common Features." *International Journal of Forecasting* 29, no. 4 (2013): 718-32.
- Hendry, D. "Granger Causality." *European journal of pure and applied mathematics* 10, no. 1 (2017).
- Hendry, D.F., and K. Hubrich. "Forecasting Aggregates by Disaggregates." *Computing in Economics and Finance* 2005 270 (2005).
- Hendry, D.F., and K. Hubrich. "Combining Disaggregate Forecasts or Combining Disaggregate Information to Forecast an Aggregate." *Journal of business and economic statistics* 29, no. 2 (2011): 216-27.
- HENDRY, David F, Massimiliano MARCELLINO, and Grayham E MIZON. "Encompassing." *Oxford Bulletin of Economics and Statistics* 70 (2008): s1.
- Hendry, D.F. "Robustifying Forecasts from Equilibrium-Correction Systems." *Journal of Econometrics* 135, no. 1 (2006): 399-426.
- Johansen, S., and B. Nielsen. "An Analysis of the Indicator Saturation Estimator as a Robust Regression Estimator." Chap. 1 In *Castle, and Shephard (2009), the Methodology and Practice*



of Econometrics: A Festschrift in Honour of David F. Hendry. . 1-36: Oxford University Press., 2009.

- Kohn, R. "When Is an Aggregate of a Time Series Efficiently Forecast by Its Past?". *Journal of Econometrics* 18, no. 3 (1982): 337-49.
- Lütkepohl, Helmut. "Forecasting Contemporaneously Aggregated Vector Arma Processes." *Journal of Business & Economic Statistics* 2, no. 3 (1984): 201-14.
- Lütkepohl, H. *Forecasting Aggregated Vector Arma Processes*. Vol. 284: Springer Verlag, 1987.
- Rose, D.E. "Forecasting Aggregates of Independent Arima Processes." *Journal of Econometrics* 5, no. 3 (1977): 323-45.
- Tiao, George C, and Irwin Guttman. "Forecasting Contemporaneous Aggregates of Multiple Time Series." *Journal of Econometrics* 12, no. 2 (1980): 219-30.
- Wei, William WS, and Bovas Abraham. "Forecasting Contemporaneous Time Series Aggregates." *Communications in Statistics-Theory and Methods* 10, no. 13 (1981): 1335-44.