

EDICIÓN E IMPUTACIÓN DE DATOS DE SERIES DE TIEMPO MÚLTIPLES: LA ESTADÍSTICA DE SACRIFICIO DE GANADO EN MÉXICO

por

Víctor M. Guerrero^{1,2} y Blanca I. Gaspar²

¹Departamento de Estadística

Instituto Tecnológico Autónomo de México (ITAM)

guerrero@itam.mx

y

²Dirección General de Contabilidad Nacional y Estadísticas Económicas

Instituto Nacional de Estadística, Geografía e Informática (INEGI)

manuel.guerrero@inegi.gob.mx y blanca.gaspar@inegi.gob.mx

**Las ideas, opiniones y comentarios expresados en este documento, son
responsabilidad exclusiva de los autores y no reflejan la posición del INEGI.**

Edición e imputación de datos de series de tiempo múltiples: la Estadística de Sacrificio de Ganado en México

por

Víctor M. Guerrero y Blanca I. Gaspar

Resumen

El tratamiento de los datos generados por la Estadística de Sacrificio de Ganado en Rastros Municipales (ESGRM) enfrenta dos problemas que se presentan comúnmente en cualquier encuesta, independientemente de lo bien diseñada y realizada que ésta sea, y que ameritan la aplicación de procedimientos estadísticos de edición e imputación de datos. El primero es cuando la unidad de información básica, el rastro municipal en el presente caso, proporciona las respuestas solicitadas, pero éstas no son congruentes de acuerdo con algún criterio empleado para validar la información. El segundo problema ocurre cuando el rastro informante deja sin responder al menos una de las preguntas importantes del cuestionario. En cualquiera de los dos casos citados, se requiere “estimar” el o los valor(es) de la(s) variable(s) con datos incongruentes o faltantes. Esta actividad se propone que sea realizada a nivel de rastro, antes de proceder a agregar los datos de un conjunto de rastros. En el caso de la ESGRM, los datos se obtienen mes a mes mediante un cuestionario que el INEGI ha diseñado para tal fin. En ese cuestionario se informa acerca de la matanza de ganado para consumo humano, principalmente de cuatro especies de ganado: bovino, porcino, ovino y caprino.

La propuesta metodológica consiste en lo siguiente. Primero se intenta obtener los datos directamente del informante, de manera que, más que estimar los datos faltantes o incongruentes, se trata de corregirlos o de conseguirlos a partir de la fuente básica. Posteriormente, la información de todos los rastros se somete a un proceso de edición que permite corregir los datos incongruentes. Luego se procede a imputar los datos faltantes con un método estadístico automatizado que hace uso de toda la información histórica disponible, tanto de la variable en cuestión como de otras variables asociadas con ella. En el presente proyecto, se eligieron del cuestionario de la ESGRM tres variables básicas: (1) número de cabezas (número de animales vivos que se introducen al rastro para su matanza), (2) peso en pie (peso del animal a su ingreso al rastro) y (3) peso en canal (peso de la carne del animal sacrificado después de retirarle algunas partes, como la piel, la cabeza y las vísceras).

La imputación de datos se propone que se realice con el método de pronósticos restringidos, basado en modelos de series de tiempo múltiples, que proporciona estimaciones óptimas, en un sentido estadístico formal. Para esto, hay necesidad de construir modelos a partir de los datos observados, utilizando técnicas estándar para este fin. Los datos empleados en la ilustración metodológica corresponden a los rastros municipales del Estado de Aguascalientes del periodo 1998-2003, en lo que toca a la modelación, mientras que los datos del año 2004 se usaron para poner a prueba la metodología en tiempo real. Estos resultados, junto con algunos experimentos de simulación, permiten confirmar la eficacia del método en la práctica. Por otro lado, queda claro que el método es viable para ser utilizado de manera rutinaria y masiva, al implementarse como parte de un sistema de cómputo.

1. INTRODUCCIÓN

Una de las tareas más importantes que realizan las Oficinas Nacionales de Estadística, como es el caso del Instituto Nacional de Estadística, Geografía e Informática (INEGI) en México, es la obtención de datos estadísticos mediante la aplicación de cuestionarios. Con esos datos se pueden medir, entre otras cosas, la situación socioeconómica y el bienestar de las poblaciones de los respectivos países. Es claro que el conocimiento de la situación que prevalece en un determinado país puede ayudar a los tomadores de decisiones, y particularmente al gobierno, a guiar sus políticas y a tomar decisiones adecuadas. Desde luego, mientras mejor sea la calidad de los datos con los que se cuente, mejores podrán ser las decisiones que se tomen.

Desde un punto de vista técnico, el conjunto de pasos que requiere el diseño de una encuesta y del cuestionario que se desea aplicar, debe hacer uso del conocimiento del fenómeno que se desea medir, así como de herramientas estadísticas relativamente complejas para convertir los datos en información útil para el tomador de decisiones. Sin embargo, la calidad de los datos que se obtengan dependerá fundamentalmente de la participación de los informantes dentro de las unidades básicas de información. Estas unidades pueden ser hogares (por ejemplo al recolectar datos sobre ingresos y gastos familiares) o establecimientos en general. En el caso de la Estadística de Sacrificio de Ganado en Rastros Municipales (ESGRM) que elabora el INEGI, ocurren los problemas típicos de datos incongruentes o faltantes, que se presentan en cualquier encuesta, independientemente de lo bien diseñada y realizada que ésta sea. La presencia de estos problemas amerita la aplicación de procedimientos estadísticos especializados para extraer información útil de los datos recolectados.

Si la cooperación del informante es pobre, se incrementa el costo de obtener datos debido a los intentos que deben repetirse para solicitar la respuesta del cuestionario. En general, la falta de cooperación del informante daña la calidad de la información estadística recopilada. Es por ello que, en principio, se debe crear conciencia en los informantes de la importancia que tiene el proporcionar datos fidedignos sobre los asuntos acerca de los que se les requiera responder. Aun cuando en México el INEGI hace un gran esfuerzo en este sentido, y se apoya en la Ley de Información Estadística y Geográfica vigente para lograr el éxito, es un hecho que la calidad de algunas cifras estadísticas todavía deja mucho que desear. En específico, la ESGRM presenta los problemas de (1) incongruencia de los datos recabados (cuando el

rastro municipal respondió el cuestionario, pero las respuestas no resultan congruentes, de acuerdo con algún criterio empleado para validar la información) o (2) datos faltantes (cuando el rastro informante dejó sin responder al menos una de las preguntas importantes del cuestionario).

La presencia de dichos problemas amerita la aplicación de procedimientos estadísticos de edición e imputación de datos, ya que en cualquiera de los dos casos citados, se requiere “estimar” el o los valor(es) de la(s) variable(s) con datos incongruentes o faltantes. Esta actividad se debe realizar a nivel de rastro, antes de agregar los datos con los de otros establecimientos del mismo tipo, para evitar la acumulación de errores. En el caso de la ESGRM, los datos se obtienen mensualmente mediante un cuestionario que el INEGI ha diseñado para tal fin (véase INEGI, 2004a). En ese cuestionario, el rastro municipal informa acerca de la matanza de ganado para consumo humano, principalmente para cuatro especies de ganado: bovino, porcino, ovino y caprino. Es deseable entonces contar con procedimientos de edición e imputación que puedan automatizarse de manera computacional para su aplicación masiva (en todos los rastros municipales) y repetitiva (mes tras mes).

La principal aportación de este trabajo radica en una propuesta de metodología estadística que brinda una solución a los problemas de edición e imputación de datos mediante el uso de modelos de series de tiempo múltiples. Con estos modelos se tiene en cuenta la información histórica de las variables en estudio, así como sus posibles interrelaciones. En el presente proyecto y a manera de ilustración del procedimiento propuesto, se eligieron del cuestionario de la ESGRM solamente tres variables básicas: (1) número de cabezas (número de animales vivos que se introducen al rastro para su matanza), (2) peso en pie (peso del animal a su ingreso al rastro) y (3) peso en canal (peso de la carne del animal sacrificado después de retirarle algunas partes, como la piel, la cabeza y las vísceras). Los datos empleados en la ilustración metodológica corresponden a los rastros municipales del Estado de Aguascalientes.

Conviene mencionar que ya han aparecido en la literatura diversos trabajos en donde se utilizan modelos de series de tiempo para sustentar procedimientos de edición e imputación de datos, tanto para el caso de una sola variable, véase al respecto Caporello y Maravall (2002) y las referencias ahí citadas; así como en el caso multivariado (Pfefferman y Nathan, 2002). También debe notarse que el procedimiento aquí propuesto difiere de lo sugerido por Caporello y Maravall, ya que aquí se trabaja

con series múltiples y, por otro lado, es más sencillo de utilizar que el método sugerido por Pfefferman y Nathan.

El contenido de este documento es el siguiente. En la Sección 2 se hace una breve descripción de la ESGRM y del esquema de estudio seguido para desarrollar este trabajo. Asimismo se presentan algunas gráficas que muestran los comportamientos de las variables en estudio, junto con sus respectivas transformaciones, las cuales sirven fundamentalmente para editar los datos de las series. En la Sección 3 se describe el modelo VAR que será empleado, junto con la metodología de pronósticos restringidos que se propone utilizar para realizar la imputación de los datos faltantes, así como otros aspectos de carácter técnico-estadístico. La Sección 4 se dedica a mostrar detalles de la construcción del modelo y los resultados de la estimación del mismo para los distintos rastros y especies considerados. En la Sección 5 se presentan algunos resultados que produce la metodología, tanto para la parte de edición como la de imputación de los datos de la ESGRM. En esa misma sección se muestran los resultados de algunas simulaciones realizadas para verificar la eficacia del método. Por último, la Sección 6 concluye este trabajo con algunas consideraciones y recomendaciones.

2. ANÁLISIS PRELIMINAR DE LOS DATOS DE LA ESGRM

La ESGRM tiene periodicidad mensual y su cobertura es nacional, pues cubre los 31 estados que conforman la República Mexicana. Se excluye al Distrito Federal debido a que en la actualidad no hay rastros en funcionamiento en esta entidad federativa (en 1993 se inició el cierre de rastros en el Distrito Federal y dos años más tarde no había ya ningún rastro abierto). El estudio se realizó con los datos de los rastros del estado de Aguascalientes, el cual fue elegido por ser el primero en orden alfabético y porque ahí se encuentra la sede del INEGI. Este estado cuenta con ocho rastros, un número lo suficientemente pequeño como para facilitar la realización de pruebas de modelos y de métodos, ya que lo que se busca es desarrollar una metodología que se pueda generalizar hacia el universo de todos los rastros municipales existentes en la República Mexicana (907 rastros en diciembre de 2004).

El periodo muestral que se utilizó en este trabajo cubre los datos de enero de 1998 a diciembre de 2003, debido a que estos eran los datos históricos disponibles en la publicación anual más reciente con la que se contaba al momento de iniciar el estudio (véase INEGI, 2004b). Los comportamientos de las tres variables, para cada uno de los

ocho rastros, se representaron mediante modelos VAR (Vectores Auto-Regresivos) y se buscó una especificación genérica que proporcionara resultados razonablemente válidos, en términos estadísticos, para todos los rastros del Estado de Aguascalientes. Para generalizar dicha solución a todos los estados de la República Mexicana, sería aconsejable realizar algunas pruebas similares a las que se realizaron en los rastros de Aguascalientes y, si el modelo genérico proporciona también resultados válidos para los demás estados, entonces se contará con un modelo aplicable a todo el universo de rastros municipales del país.

2.1 Definición de conceptos

En la ESGRM se emplean los siguientes conceptos (véase INEGI, 2004a). *Rastro municipal* es la unidad básica de observación, que se define como el inmueble instalado por los gobiernos de los estados, en coordinación con los municipios, en donde se lleva a cabo la matanza de ganado para consumo humano. La *especie de ganado* se refiere a la clasificación de los animales, que en general incluye a todos los individuos con rasgos comunes entre sí y que, por lo tanto, pueden ubicarse dentro de una misma categoría. Así, se define entonces al ganado *bovino* como la especie animal doméstica cuya finalidad zootécnica, de acuerdo a su raza o hibridación, es la producción de leche, carne o ambas; incluye toros, bueyes, vacas, toretes, vaquillas y becerros. De igual forma, el ganado *porcino* es otra especie cuya finalidad zootécnica es fundamentalmente la producción de carne y manteca; también se le denomina como marrano, puerco, cochino, cerdo o chanco. El ganado *ovino* es la especie cuya finalidad, de acuerdo a su raza o hibridación, es fundamentalmente la producción de carne o lana; se refiere a los borregos, corderos y carneros. Finalmente, el ganado *caprino* es la especie animal cuya finalidad es fundamentalmente la producción de carne o leche; se les conoce como cabras o chivas.

Dada la importancia que tienen para la validación de datos los conceptos de peso en pie promedio (cociente del peso en pie entre el número de cabezas) y aprovechamiento de la carne (cociente del peso en canal entre el peso en pie), se realizó primero un análisis gráfico de estas variables, las cuales se derivan de las que originalmente se captan en el cuestionario. En dichas gráficas se observó que estas dos nuevas variables se mantienen cercanas a un valor constante por varios meses consecutivos. Las figuras 2.1 y 2.2 muestran ejemplos de este fenómeno para los datos de algunos rastros de Aguascalientes. Por razones de confidencialidad, en todas las

figuras y cuadros de este documento se omiten los nombres de los rastros. La especie de ganado en consideración en la Figura 2.1 es bovinos y en la Figura 2.2 es porcinos.

Figura 2.1 Peso promedio y aprovechamiento de la carne de bovinos (Rastro X1)

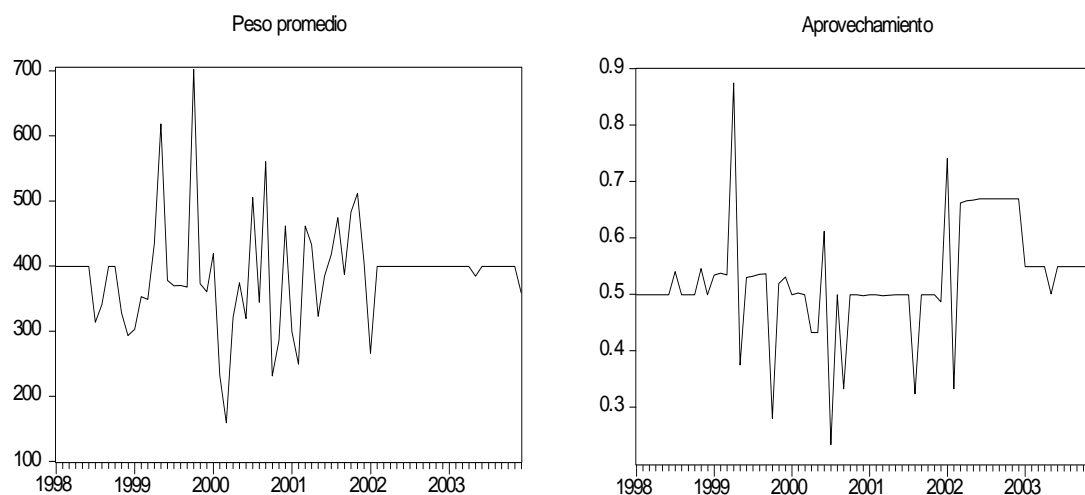
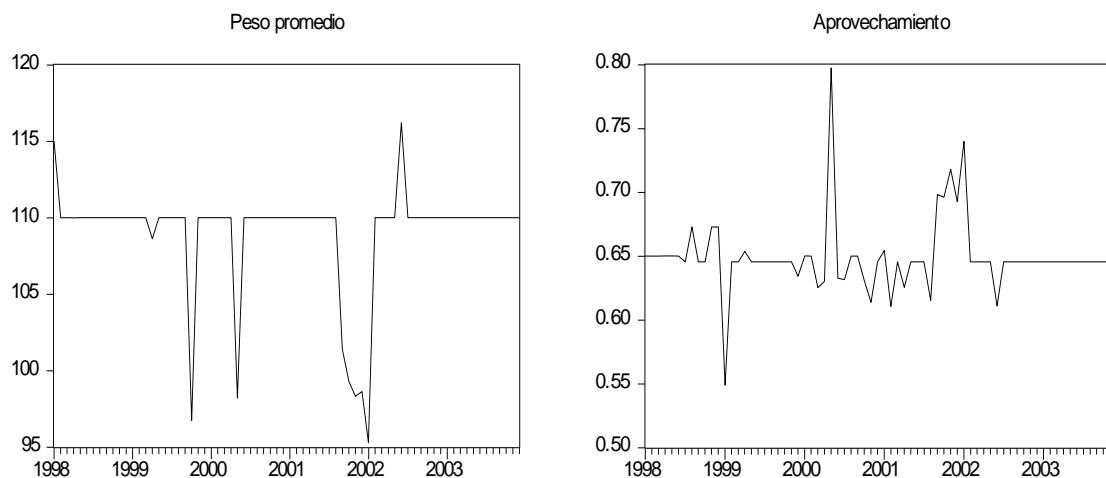


Figura 2.2 Peso promedio y aprovechamiento de la carne de porcinos (Rastro X8)



El comportamiento básico que se observa en estas dos figuras se repite para todas las especies y rastros en consideración, lo cual conduce a pensar que con mucha frecuencia, los informantes dan poca importancia al llenado del cuestionario y simplemente reportan cifras alrededor de ciertos valores “promedio” o dentro de unas determinadas bandas, que son conocidas y que permiten identificar como datos atípicos a los que se encuentran fuera de las bandas citadas. En otros casos, los datos también suelen reportarse con errores evidentes, que suelen editarse por el personal encargado de

la captura de información en el INEGI. Aun con esta primera validación de la información, ocurre que algunos de estos errores pasan desapercibidos por el personal de captura, o estas personas, al hacer la captura de la información, introducen nuevos errores en los datos. Para corregir de manera sistemática este tipo de errores, lo que se propone es considerar los límites y las características permisibles de las variables; por ejemplo, se sabe que el dato del número de cabezas debe ser entero y no-negativo.

Así pues, aunque las cifras del documento del INEGI (2004b) son de carácter definitivo, fue necesario someterlas a un análisis previo que asegurara que los datos cumplieran con los límites permisibles para la información. De esta forma se obtuvieron los datos editados de las series, que se considerarían después para la construcción del modelo que se describe más adelante. Una vez realizada la edición preliminar de los datos, surgió la necesidad de considerar una transformación que permitiera a los datos mostrar su variabilidad natural, sin estar sujetos a mantenerse dentro de ciertos límites. Lo anterior se hizo después de observar en el análisis gráfico que el peso en pie promedio y el aprovechamiento de la carne, seguían comportamientos prácticamente constantes por varios periodos consecutivos. Este procedimiento, que fue aplicado para corregir la serie histórica, se debe aplicar igualmente con cada mes adicional de información que se considere. Es decir, para los datos de un nuevo mes, primero se debe realizar una inspección que permita revisar si los datos de las variables guardan congruencia entre sí, y una vez pasada esta etapa se les aplicará una transformación. Los nuevos datos, ya transformados, pasarán entonces al procedimiento siguiente en donde se imputarán los valores faltantes mediante el uso del modelo y de la técnica de pronósticos restringidos.

2.2 Transformaciones de las variables

Para aprovechar el conocimiento del patrón que siguen las series de tiempo en estudio y mejorar la estimación de los datos faltantes o incongruentes, se decidió transformar los datos que originalmente se captan en los cuestionarios. Las transformaciones propuestas, que se describen a continuación, permiten complementar la edición de los datos para que satisfagan los criterios numéricos que los informantes han tratado de aplicar rutinariamente, aunque de manera informal. Por ello se considera que la aplicación de las transformaciones constituye una parte fundamental del procedimiento de edición de los datos.

Las variables en estudio se definen a continuación. Para el mes t , NC_t denota el número de cabezas de ganado, PP_t el peso en pie, medido en kilogramos, y PC_t el peso

en canal, también medido en kilogramos. Para transformar la variable NC se usó el hecho de que $0 \leq NC_t$ para $t = 1, \dots, N$, por lo cual, $0 < NC_t + 0.1$, donde la constante 0.1 se eligió para volver positivos los valores 0 de esta variable, sin afectar prácticamente la dinámica de la serie, debido a que es un valor pequeño en términos relativos (ya que la variable se mide en unidades enteras y 1/10 de cabeza de ganado no es confundible con un dato observado de la serie). De esta manera se sabe que

$$0 < (NC_t + 0.1) / (NC_{t-1} + 0.1) < \infty \quad (2.1)$$

y, por lo tanto, la transformación propuesta es

$$-\infty < TNC_t = \log\left(\frac{NC_t + 0.1}{NC_{t-1} + 0.1}\right) < \infty. \quad (2.2)$$

Así, la variable en diferencias que aparecerá en el modelo VAR será

$$\begin{aligned} DTNC_t &= \log\left(\frac{NC_t + 0.1}{NC_{t-1} + 0.1}\right) - \log\left(\frac{NC_{t-1} + 0.1}{NC_{t-2} + 0.1}\right) \\ &\cong \log(NC_t / NC_{t-1}) - \log(NC_{t-1} / NC_{t-2}) \\ &= \log(1 + r_t^{NC}) - \log(1 + r_{t-1}^{NC}) \\ &\cong r_t^{NC} - r_{t-1}^{NC} \end{aligned} \quad (2.3)$$

con r_t^{NC} la tasa de crecimiento relativo de NC_t . La razón para trabajar con la variable en diferencias se dará posteriormente, cuando se haga referencia a la construcción del modelo.

En lo que respecta a la variable PP, se sabe que para cada momento de observación t , el peso en pie promedio debe cumplir con restricciones del tipo $K1 \leq PP_t / NC_t \leq K2$ con $K1$ y $K2$ constantes positivas conocidas (por ejemplo, para el ganado bovino $K1 = 250$ y $K2 = 550$), de manera que se deben cumplir las desigualdades $K1 < PP_t / NC_t + 0.1$ y $PP_t / NC_t - 0.1 < K2$, donde la constante 0.1 se eligió de manera similar al caso anterior. Aunque en el presente caso, las unidades de 0.1 corresponden a las de la variable PP/NC , que se mide en kilogramos, y el efecto de la inclusión de esta constante es el de ampliar el intervalo $[K1, K2]$ en 100 gramos de cada lado, para trabajar entonces con el intervalo $[K1 - 0.1, K2 + 0.1]$. En consecuencia, se obtiene

$$0 < (PP_t / NC_t + 0.1 - K1) / (K2 - PP_t / NC_t + 0.1) < \infty \quad (2.4)$$

y la transformación a usar es ahora

$$-\infty < TPP_t = \log\left(\frac{PP_t / NC_t + 0.1 - K1}{K2 - PP_t / NC_t + 0.1}\right) < \infty. \quad (2.5)$$

Así pues, la variable que se usará en el modelo resulta ser

$$\begin{aligned}
DTPP_t &= \log\left(\frac{PP_t / NC_t + 0.1 - K1}{K2 - PP_t / NC_t + 0.1}\right) - \log\left(\frac{PP_{t-1} / NC_{t-1} + 0.1 - K1}{K2 - PP_{t-1} / NC_{t-1} + 0.1}\right) \\
&= \log\left(\frac{PP_t / NC_t + 0.1 - K1}{K2 - PP_t / NC_t + 0.1} \bigg/ \frac{PP_{t-1} / NC_{t-1} + 0.1 - K1}{K2 - PP_{t-1} / NC_{t-1} + 0.1}\right) \\
&= \log\left(\frac{PP_t / NC_t + 0.1 - K1}{PP_{t-1} / NC_{t-1} + 0.1 - K1} \bigg/ \frac{K2 - PP_t / NC_t + 0.1}{K2 - PP_{t-1} / NC_{t-1} + 0.1}\right) \\
&\cong \log\left(\frac{PP_t / NC_t - K1}{PP_{t-1} / NC_{t-1} - K1}\right) - \log\left(\frac{K2 - PP_t / NC_t}{K2 - PP_{t-1} / NC_{t-1}}\right) \\
&= \log(1 + r_t^{PP/NC-K1}) - \log(1 + r_t^{K2-PP/NC}) \\
&\cong r_t^{PP/NC-K1} - r_t^{K2-PP/NC}
\end{aligned} \tag{2.6}$$

Por lo que toca a la variable PC, se sabe que el aprovechamiento de la carne debe satisfacer las restricciones $K3 \leq PC_t/PP_t \leq K4$ para toda t , con $K3$ y $K4$ constantes positivas conocidas (por ejemplo, para el ganado bovino $K3 = 0.5$ y $K4 = 0.55$). Para poner a PC en relación directa con NC (que comúnmente no tiene datos faltantes) y no con PP (que con frecuencia tiene datos faltantes) se hará uso del siguiente argumento

$$K1(PC_t / PP_t) \leq (PP_t / NC_t)(PC_t / PP_t) \leq K2(PC_t / PP_t) \tag{2.7}$$

de tal forma que se obtiene $K1 \cdot K3 \leq PC_t/NC_t \leq K2 \cdot K4$. Por lo tanto, de manera similar al caso anterior, ahora se usa la transformación

$$-\infty < TPC_t = \log\left(\frac{PC_t / NC_t + 0.01 - K1 \cdot K3}{K2 \cdot K4 - PC_t / NC_t + 0.01}\right) < \infty \tag{2.8}$$

donde la constante elegida para garantizar las desigualdades, sin distorsionar la dinámica de la serie, es 0.01. Esto implica que el intervalo $[K1 \cdot K3, K2 \cdot K4]$ se amplía en 10 gramos de cada lado, ya que las unidades de PC/NC son kilogramos. Así pues, la variable que se usará en el modelo es

$$DTPC_t \cong r_t^{PC/NC-K1 \cdot K3} - r_t^{K2 \cdot K4-PC/NC} \tag{2.9}$$

Las gráficas de las figuras 2.3 a 2.6, que se presentan a continuación, muestran el comportamiento de las series de tiempo para las variables en la escala original (NC, PP y PC), para las variables transformadas (TNC, TPP y TPC), y para las primeras diferencias de las series transformadas (DTNC, DTPP y DTPC), de algunos rastros en estudio y de las cuatro especies de ganado.

Figura 2.3 Rastro X3
(Especie Bovinos)

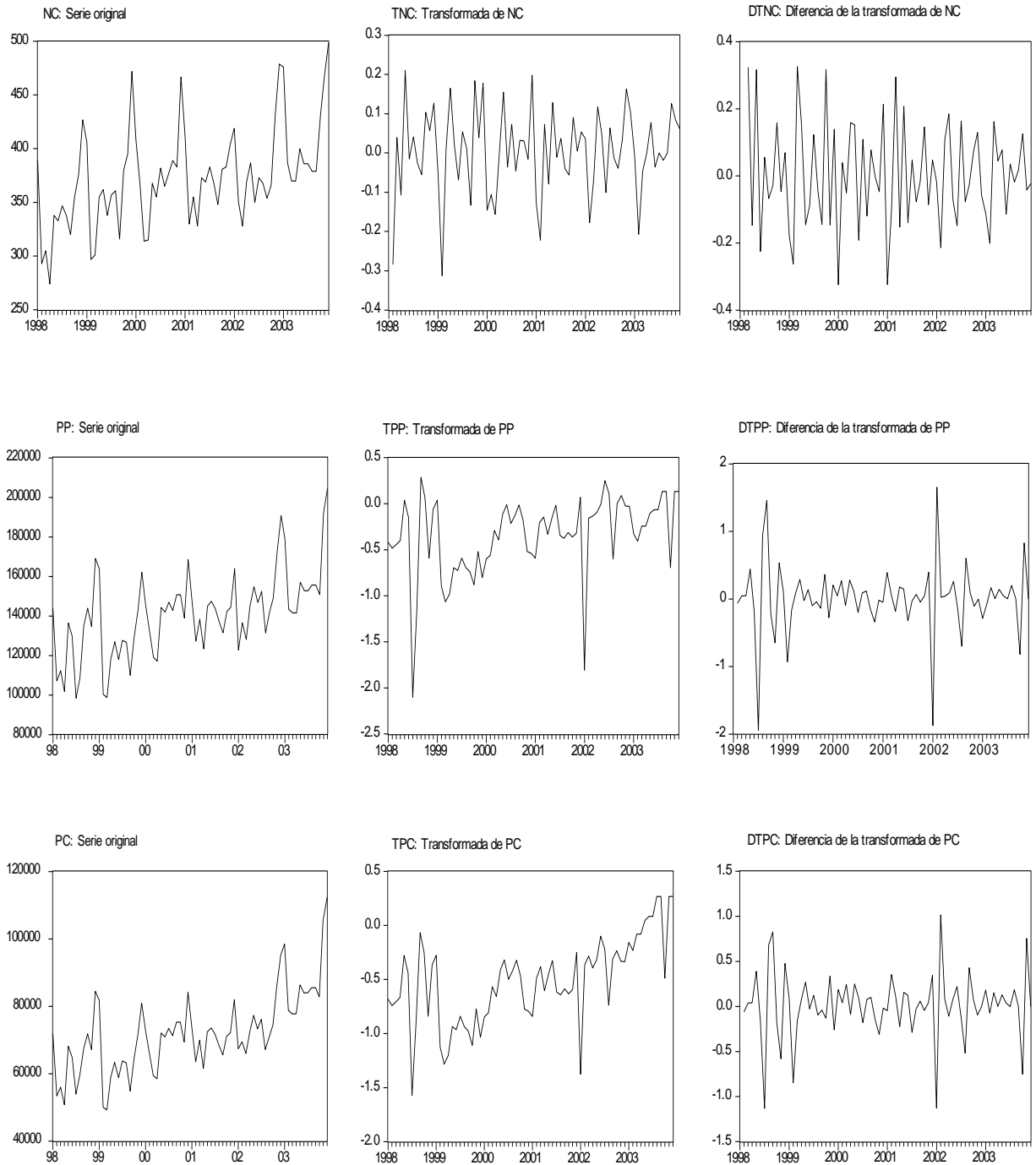


Figura 2.4 Rastro X3
(Especie Porcinos)

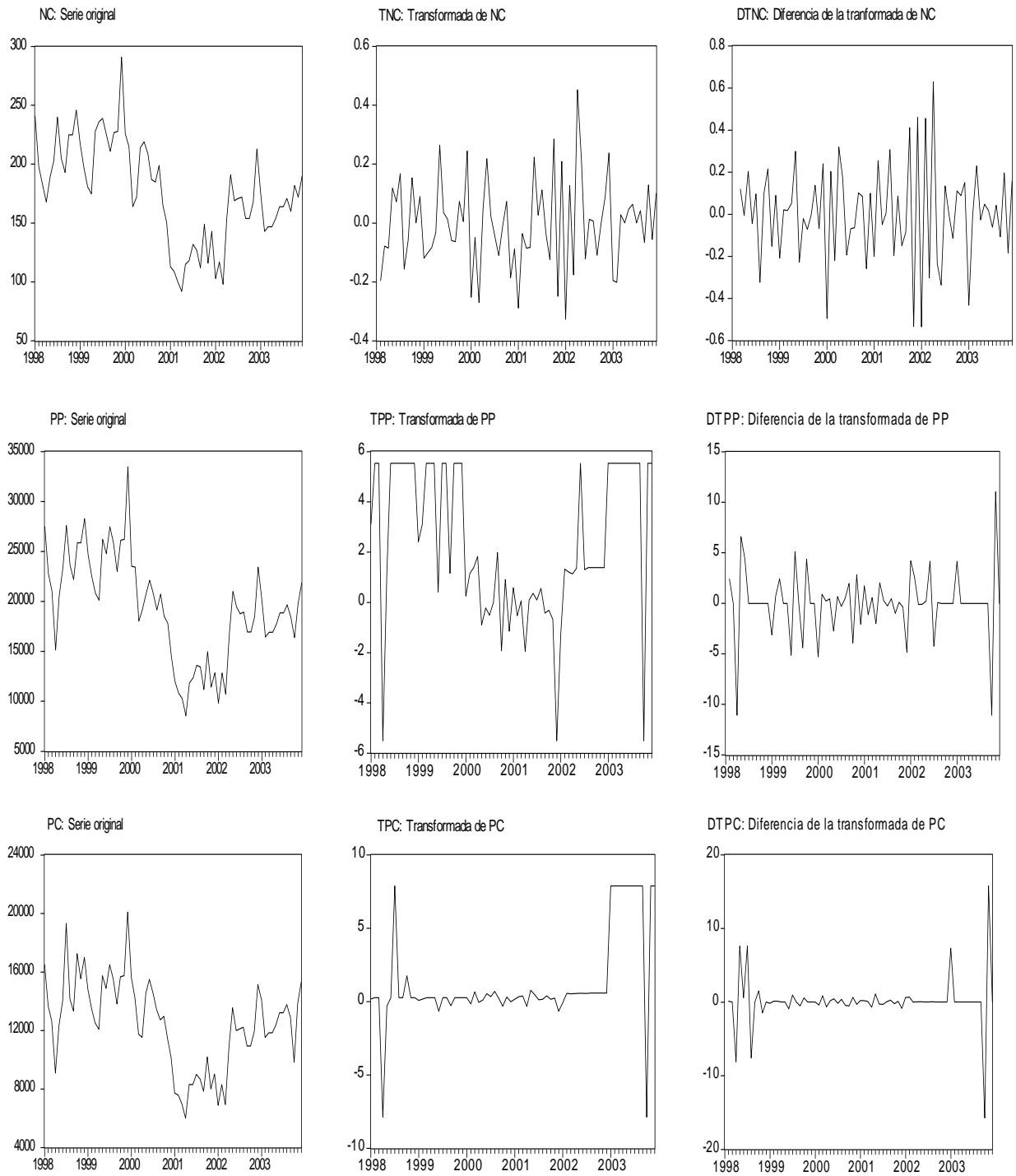


Figura 2.5 Rastro X8
(Especie Ovinos)

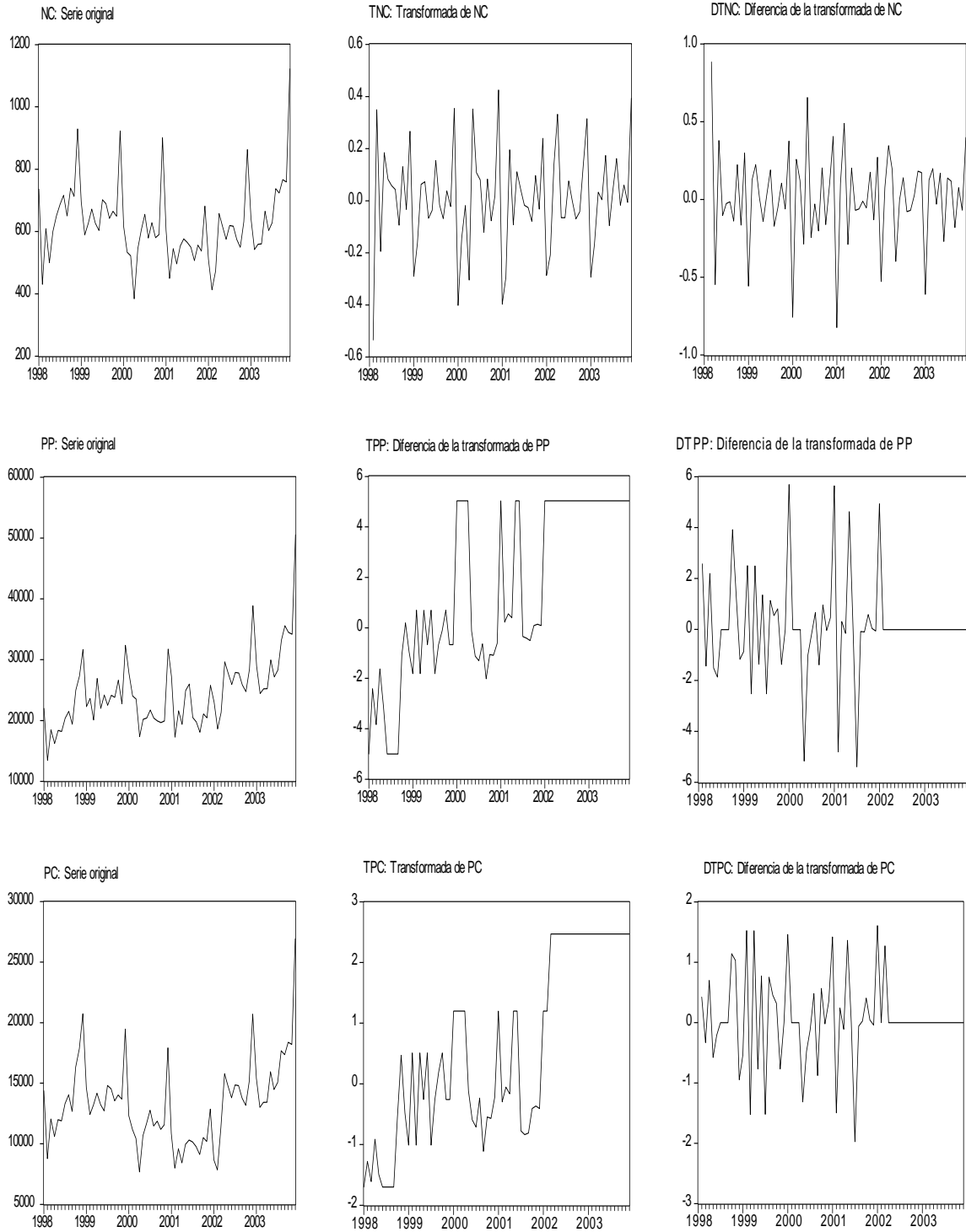
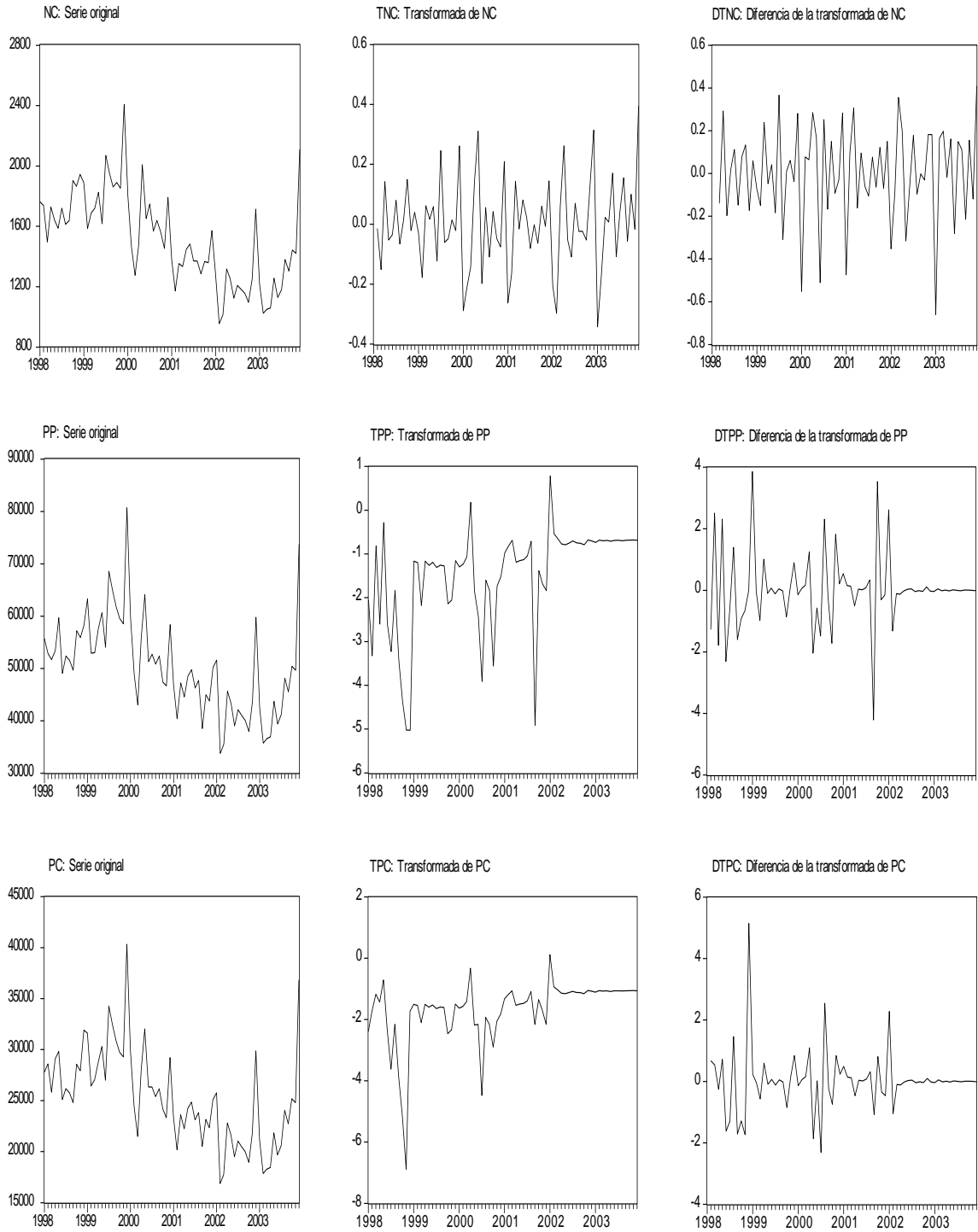


Figura 2.6 Rastro X8
(Especie Caprinos)



Las constantes K1, K2, K3 y K4 que se emplean en las transformaciones de los datos de la ESGRM surgen por la necesidad de contar con intervalos en donde se ubiquen típicamente las cifras observadas del peso en pie promedio y del aprovechamiento de la carne por especie (según se considera en Ruíz *et al.*, 2001). Dada la gran variedad de pesos por especie y, dentro de cada especie, por raza y por edad, dichos valores se obtuvieron en principio, de las páginas disponibles en Internet, de la Facultad de Veterinaria de la Universidad Nacional Autónoma de México (Veterinaria, 2005, www.veterin.unam.mx), de la Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación (SAGARPA, 2005, www.sagarpa.gob.mx/ganaderito), y de asociaciones ganaderas (Ganaderos, 2005, www.mexicoganadero.com/limousin). Finalmente, la unificación de las diversas constantes encontradas se realizó a través de consultas directas con el personal veterinario que labora en los rastros municipales.

3. METODOLOGÍA ESTADÍSTICA

El aspecto metodológico más importante de este trabajo se refiere a la técnica estadística de pronósticos restringidos, la cual se basa en modelos estadísticos que representan a la serie de tiempo múltiple en estudio. Estos modelos constituyen una herramienta muy útil para capturar la dinámica de las variables que conforman el vector de series de tiempo. Aquí se hace uso de uno de tales modelos, que se expresa de manera relativamente sencilla y que resulta muy apropiado para efectuar pronósticos. De hecho, se enfatiza la utilidad del modelo con fines de pronóstico y se muestra el desarrollo de un método de imputación de datos, basado en la idea de los pronósticos restringidos.

3.1 Modelo VAR

Considérese un vector $\mathbf{Z}_t = (Z_{1t}, \dots, Z_{kt})'$ que contiene las observaciones de k series de tiempo, para $t = 1, \dots, N$. La representación en forma de Vector Auto-Regresivo (VAR) corresponde al modelo

$$\Pi(B)\mathbf{Z}_t = \Lambda\mathbf{D}_t + \mathbf{a}_t \quad (3.1)$$

en donde $\Pi(B)$ es un polinomio matricial de orden $p < \infty$, en el operador de retraso B , tal que $B\mathbf{Z}_t = \mathbf{Z}_{t-1}$ para cada \mathbf{Z} y t , o sea,

$$\Pi(B) = \mathbf{I}_k - \Pi_1 B - \dots - \Pi_p B^p \quad (3.2)$$

con I_k la matriz identidad de orden k y

$$\Pi_j = \begin{pmatrix} \pi_{j,11} & \pi_{j,12} & \dots & \pi_{j,1k} \\ \pi_{j,21} & \pi_{j,22} & \dots & \pi_{j,2k} \\ \dots & \dots & \dots & \dots \\ \pi_{j,k1} & \pi_{j,k2} & \dots & \pi_{j,kk} \end{pmatrix} \text{ para } j = 1, \dots, p. \quad (3.3)$$

El modelo incluye el vector $\mathbf{D}_t = (D_{1t}, \dots, D_{kt})'$ que contiene variables determinísticas (para representar los niveles de las series, así como efectos de estacionalidad o de posibles intervenciones), cuya influencia sobre \mathbf{Z}_t se captura mediante la matriz de parámetros Λ . La sucesión de vectores $\{\mathbf{a}_t\}$ se considera generada por un proceso de Ruido Blanco Gaussiano con media cero, o sea, está formada por vectores aleatorios independientes e idénticamente distribuidos, con distribución Normal, o sea

$$\mathbf{a}_t \sim N_k(\mathbf{0}_k, \Sigma_a) \text{ para } t = 1, \dots, N \quad (3.4)$$

donde $\mathbf{0}_k$ es el vector cero de dimensión k y Σ_a es la matriz de varianza – covarianza contemporánea de los errores, que está dada por

$$\Sigma_a = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1k} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2k} \\ \dots & \dots & \dots & \dots \\ \sigma_{k1} & \sigma_{k2} & \dots & \sigma_k^2 \end{pmatrix} \quad (3.5)$$

en donde $\sigma_{ij} = \text{Cov}(a_{it}, a_{jt})$ para $i \neq j = 1, \dots, k$ y $\sigma_i^2 = \text{Var}(a_{it})$ para $i = 1, \dots, k$.

Un modelo VAR de orden finito puede concebirse de diversas maneras: (i) como la forma reducida de un modelo econométrico estructural; (ii) como una aproximación a la representación autorregresiva pura de un modelo del tipo Vector Auto-Regresivo de Promedios Móviles (VARMA) invertible; o (iii) como una representación útil para capturar las regularidades empíricas observadas en la serie vectorial $\{\mathbf{Z}_t\}$. Esta última interpretación del modelo VAR es básicamente la que explotó Sims (1980) para solucionar el problema de la construcción de modelos econométricos estructurales, cuando la teoría económica no está lo suficientemente explícita o completamente desarrollada para este fin. Una excelente referencia acerca de los modelos VARMA la constituye el texto de Lütkepohl (1991).

Si las raíces de la ecuación determinante $|\Pi(x)| = 0$ se encuentran fuera del círculo unitario, entonces la serie $\{\mathbf{Z}_t\}$ resulta ser estacionaria y el modelo (3.1) está bien definido, de otra forma la serie no es estacionaria y será necesario suponer que el

proceso comenzó en un punto finito del tiempo, con condiciones iniciales fijas, para que el modelo (3.1) quede bien definido. Cuando las variables del vector de series de tiempo son todas integradas de orden uno y no están cointegradas, se podría aplicar el operador diferencia $\nabla = 1 - B$ a todas las variables del modelo, con lo cual se obtendría una representación apropiada para $\{\nabla \mathbf{Z}_t\}$ y no para la serie $\{\mathbf{Z}_t\}$. En cambio, si existe alguna relación de cointegración entre las variables (véase Engle y Granger, 1987), se podría trabajar con el siguiente modelo VAR en forma de Corrección de Errores (VEC)

$$\Pi^*(B)\nabla \mathbf{Z}_t = \Lambda \mathbf{D}_t - \Pi(1)\mathbf{Z}_{t-1} + \mathbf{a}_t. \quad (3.6)$$

que surge de la igualdad $\Pi(B) = \Pi(1)B + \Pi^*(B)\nabla$. En este caso, el polinomio matricial $\Pi^*(B)$ es de orden $p-1$, mientras que la matriz $\Pi(1)$ está definida de acuerdo con las relaciones de cointegración que existan entre las variables.

Nótese en la ecuación (3.6) que $\Pi^*(B)\nabla \mathbf{Z}_t$ captura las relaciones de corto plazo de \mathbf{Z}_t , mientras que $\Pi(1)\mathbf{Z}_{t-1}$ representa las relaciones de largo plazo. Además, todos los elementos de dicha ecuación son estacionarios, por lo cual la ecuación $|\Pi^*(x)| = 0$ tiene todas sus raíces fuera del círculo unitario. Conviene resaltar que el modelo VEC está asociado con teoría económica, ya que permite interpretar los resultados y realizar inferencias acerca de las relaciones económicas, tanto de corto como de largo plazo. Por otro lado, es importante hacer notar que la información capturada por los modelos VAR y VEC es exactamente la misma, aunque las representaciones (3.1) y (3.6) sean formalmente distintas. El otro uso del modelo, que resulta de primordial importancia para este trabajo, es el de pronóstico. En el caso del presente estudio, el pronóstico que se desea obtener será de un solo periodo por delante, por lo cual no se hará uso de las relaciones de largo plazo en ningún momento. Así pues, en lo que sigue se utilizará la forma VAR (que no requiere del análisis de cointegración) y la metodología de pronósticos restringidos se presentará para ese modelo en especial. En Guerrero, *et al.* (2005) se muestra la correspondiente metodología que es aplicable con modelos VEC.

3.2 Pronósticos restringidos

Para desarrollar la metodología de pronósticos con restricciones, se supondrá que el modelo y sus parámetros son conocidos, de manera tal que, por ahora, no se entrará en consideraciones respecto a la especificación, estimación, ni validación del modelo, aunque en la aplicación práctica sí se deben considerar tales aspectos. El vector de información $\mathbf{Z} = (\mathbf{Z}_1', \dots, \mathbf{Z}_N')'$ contiene las observaciones de $\mathbf{Z}_t = (\text{TNC}_t, \text{TPP}_t, \text{TPC}_t)'$ para $t = 1, \dots, N$, de forma que contiene todos los datos disponibles de la serie

múltiple en estudio, mientras que \mathbf{Z}_{N+1} es el vector de valores futuros que se desea pronosticar, a partir del origen ubicado en el tiempo N . Debido a que las series estacionarias son $\nabla \mathbf{Z}_t = (\text{DTNC}_t, \text{DTPP}_t, \text{DTPC}_t)'$ y que no existe interés en realizar análisis de cointegración, el modelo VAR por utilizar es

$$\Pi(B)\nabla \mathbf{Z}_t = \Lambda \mathbf{D}_t + \mathbf{a}_t. \quad (3.7)$$

De esta manera, el pronóstico lineal óptimo de $\nabla \mathbf{Z}_{N+1}$ en el sentido de ECM mínimo, está dado por la esperanza condicional dada toda la información histórica disponible, o sea,

$$E(\nabla \mathbf{Z}_{N+1} | \mathbf{Z}) = \Lambda \mathbf{D}_{N+1} + \Pi_1 E(\nabla \mathbf{Z}_N | \mathbf{Z}) + \dots + \Pi_p E(\nabla \mathbf{Z}_{N+1-p} | \mathbf{Z}). \quad (3.8)$$

Este pronóstico produce el vector de errores de pronóstico

$$\nabla \mathbf{Z}_{N+1} - E(\nabla \mathbf{Z}_{N+1} | \mathbf{Z}) = \mathbf{a}_{N+1} \quad (3.9)$$

cuya correspondiente matriz de Error Cuadrático Medio (ECM) es

$$\text{ECM}[E(\nabla \mathbf{Z}_{N+1} | \mathbf{Z})] = \text{Var}(\mathbf{a}_{N+1} | \mathbf{Z}) = \Sigma_a. \quad (3.10)$$

Supóngase ahora que se conoce también el vector de observaciones $\mathbf{Y} = (Y_1, \dots, Y_M)'$ que impone un conjunto de $M \geq 0$ restricciones linealmente independientes sobre los valores futuros del vector de variables \mathbf{Z} . Dichas restricciones provienen de una fuente externa al modelo de series de tiempo, y el vector de observaciones guarda la siguiente relación con el vector de valores futuros

$$\mathbf{Y} = C \nabla \mathbf{Z}_{N+1} + \mathbf{u} \quad (3.11)$$

donde $\mathbf{u} = (u_1, \dots, u_M)'$ es un vector aleatorio que se distribuye como $N(\mathbf{0}_M, \Sigma_u)$. La matriz C es de dimensión $M \times k$, se supone conocida y de rango $M \leq k$. Más adelante se mostrará la forma que toma C para algunos casos específicos del problema de imputación que se presenta en la ESGRM, en el cual los datos Y_1, \dots, Y_M corresponden a los valores de las variables que son conocidos al tiempo $N+1$.

Debe notarse que la expresión (3.11) corresponde a un conjunto de restricciones lineales estocásticas que se desea imponer sobre $\nabla \mathbf{Z}_{N+1}$. También se supondrá que $E(\mathbf{u} | \mathbf{Z}) = \mathbf{0}$, de tal forma que las restricciones son insesgadas, pero tienen incertidumbre asociada con la varianza $\text{Var}(\mathbf{u} | \mathbf{Z}) = \Sigma_u$. Si además se cumple que $E(\mathbf{a}_{N+1} \mathbf{u}' | \mathbf{Z}) = 0$, entonces se obtiene el pronóstico restringido óptimo de $\nabla \mathbf{Z}_{N+1}$ dada la serie histórica \mathbf{Z} y las restricciones \mathbf{Y} , esto es

$$\nabla \mathbf{Z}_{N+1} = E(\nabla \mathbf{Z}_{N+1} | \mathbf{Z}) + A_u [\mathbf{Y} - C E(\nabla \mathbf{Z}_{N+1} | \mathbf{Z})] \quad (3.12)$$

con

$$A_u = \Sigma_a C' (C \Sigma_a C' + \Sigma_u)^{-1} \quad (3.13)$$

y

$$ECM(\nabla \hat{\mathbf{Z}}_{N+1}) = (I_k - A_u C) \Sigma_a. \quad (3.14)$$

En el Apéndice se demuestra este resultado, aunque una demostración del caso más general, en el cual se pronostican $H > 1$ periodos por delante, se encuentra en Pankratz (1989). Además, si no hay incertidumbre en la información adicional \mathbf{Y} , se tiene que $\Sigma_u = 0$, de manera que el resultado anterior se convierte en

$$\nabla \hat{\mathbf{Z}}_{N+1} = E(\nabla \mathbf{Z}_{N+1} | \mathbf{Z}) + A [\mathbf{Y} - C E(\nabla \mathbf{Z}_{N+1} | \mathbf{Z})] \quad (3.15)$$

con

$$A = \Sigma_a C' (C \Sigma_a C')^{-1} \quad (3.16)$$

y

$$ECM(\nabla \hat{\mathbf{Z}}_{N+1}) = (I_k - AC) \Sigma_a. \quad (3.17)$$

Este último resultado es de particular importancia para la aplicación a la ESGRM, puesto que los datos para el tiempo $N+1$ se supondrá que no están sujetos a incertidumbre. También es de subrayar que la combinación de información que se logra es óptima, en tanto que el pronóstico restringido $\nabla \hat{\mathbf{Z}}_{N+1}$ tiene ECM mínimo dentro de la clase de pronósticos lineales. Además, el pronóstico restringido tiene menor ECM que el pronóstico irrestricto, ya que se cumple la igualdad

$$ECM[E(\nabla \mathbf{Z}_{N+1} | \mathbf{Z})] = ECM(\nabla \hat{\mathbf{Z}}_{N+1}) + AC \Sigma_a \quad (3.18)$$

en donde $AC \Sigma_a = \Sigma_a C' (C \Sigma_a C')^{-1} C \Sigma_a$ es una matriz semidefinida positiva.

3.3 Pronósticos de las variables originales

Debe notarse que los pronósticos restringidos se obtienen para las series en diferencias de forma que, para traerlos a las unidades de las variables originales, los pronósticos deben expresarse primero en las escalas de los niveles de las series transformadas y después retransformarse con las transformaciones inversas $T^{-1}(\cdot)$ aplicables a cada una de las variables. Para obtener los pronósticos de las variables NC_{N+1} , PP_{N+1} y PC_{N+1} , a partir de los pronósticos del vector $\nabla \mathbf{Z}_{N+1} = (DTNC_{N+1}, DTPP_{N+1}, DTPC_{N+1})'$, primero se hace uso de las expresiones

$$TNC_{N+1} = TNC_N + DTNC_{N+1}, \quad TPP_{N+1} = TPP_N + DTPP_{N+1}$$

y

$$(3.19)$$

$$TPC_{N+1} = TPC_N + DTPC_{N+1},$$

que producen los valores de las variables transformadas en niveles. Después se regresa a las unidades de las variables originales de la forma siguiente. Para el caso de NC se sabe que $\exp(TNC_{N+1}) = (NC_{N+1} + 0.1) / (NC_N + 0.1)$, de donde

$$NC_{N+1} = (NC_N + 0.1) \exp(TNC_{N+1}) - 0.1 \quad (3.20)$$

Luego, para PP se obtiene primero

$$\begin{aligned} PP_{N+1}/NC_{N+1} &= (K2 - PP_{N+1}/NC_{N+1} + 0.1) \exp(TPP_{N+1}) + K1 - 0.1 \\ &= (K2+0.1) \exp(TPP_{N+1}) + K1 - 0.1 - (PP_{N+1}/NC_{N+1}) \exp(TPP_{N+1}) \end{aligned} \quad (3.21)$$

de donde se sigue que

$$(PP_{N+1}/NC_{N+1})[1 + \exp(TPP_{N+1})] = (K2 + 0.1) \exp(TPP_{N+1}) + K1 - 0.1 \quad (3.22)$$

y, finalmente,

$$PP_{N+1} = NC_{N+1}[(K2 + 0.1) \exp(TPP_{N+1}) + K1 - 0.1] / [1 + \exp(TPP_{N+1})]. \quad (3.23)$$

En lo que toca a la variable PC, de igual manera que el resultado inmediato anterior, ahora se deduce que

$$PC_{N+1} = NC_{N+1}[(K2 \cdot K4 + 0.01) \exp(TPC_{N+1}) + K1 \cdot K3 - 0.01] / [1 + \exp(TPC_{N+1})] \quad (3.24)$$

Es conveniente notar que, si se conoce el valor de la variable PP_{N+1} , éste podría utilizarse para obtener un valor pronosticado de PC_{N+1} que satisfaga estrictamente la restricción original $K3 \leq PC_{N+1}/PP_{N+1} \leq K4$, para ello se debe usar la expresión

$$PC_{N+1} = PP_{N+1}[(K4 + 0.01) \exp(TPC_{N+1}) + K3 - 0.01] / [1 + \exp(TPC_{N+1})] \quad (3.25)$$

en lugar de la expresión anterior.

3.4 Verificación de compatibilidad

No obstante que $\nabla \hat{\mathbf{Z}}_{N+1}$ proporciona un pronóstico óptimo, debe notarse que las expresiones (3.15) – (3.17) implican combinar información. Esta combinación debe juzgarse desde un punto de vista empírico, pues pudiera ocurrir que las fuentes indicaran comportamientos contradictorios. Para que la combinación tenga validez empírica, lo razonable es verificar la compatibilidad entre los datos que provienen de las diferentes fuentes de información. Por ello se considera que la información acerca de los valores futuros proveniente de la fuente externa al modelo, \mathbf{Y} , es compatible con la correspondiente información provista por el modelo, $CE(\nabla \mathbf{Z}_{N+1}|\mathbf{Z})$, si la distancia entre estos dos vectores es cercana a cero. Con un argumento similar al expuesto en Guerrero *et al.* (2005), se define ahora el vector de diferencias

$$\mathbf{d} = \mathbf{Y} - CE(\nabla \mathbf{Z}_{N+1}|\mathbf{Z}) = \mathbf{Ca}_{N+1} \quad (3.26)$$

el cual, antes de observar los valores de \mathbf{Y} y $CE(\nabla \mathbf{Z}_{N+1}|\mathbf{Z})$, se distribuye como $N(\mathbf{0}_M, C\Sigma_a C')$. Por lo tanto, la distancia que tiene en cuenta la variabilidad en \mathbf{d} , toma la forma del estadístico

$$\mathbf{K} = \mathbf{d}'(C\Sigma_a C')^{-1}\mathbf{d} \sim \chi_M^2. \quad (3.27)$$

En consecuencia, puede afirmarse que

$$\Pr[\mathbf{K} \leq \chi_M^2(\alpha)] = 1 - \alpha \quad (3.28)$$

con $\chi_M^2(\alpha)$ el punto porcentual superior α de la distribución Ji-cuadrada con M grados de libertad, y de aquí se deduce que $\mathbf{Y} - CE(\nabla \mathbf{Z}_{N+1}|\mathbf{Z})$ se encuentra dentro de la región de compatibilidad si

$$\mathbf{K}_{\text{calc}} = [\mathbf{Y} - CE(\nabla \mathbf{Z}_{N+1}|\mathbf{Z})]'(C\Sigma_a C')^{-1}[\mathbf{Y} - CE(\nabla \mathbf{Z}_{N+1}|\mathbf{Z})] \leq \chi_M^2(\alpha). \quad (3.29)$$

Equivalentemente, \mathbf{Y} es incompatible con $CE(\nabla \mathbf{Z}_{N+1}|\mathbf{Z})$, al nivel de significancia de $100\alpha\%$, si no se cumple (3.29). Debe notarse que esta regla de decisión es válida asintóticamente, puesto que se basa en la distribución Ji-cuadrada del estadístico \mathbf{K} , para lo cual se supuso que todos los parámetros involucrados en el modelo, incluyendo Σ_a , son conocidos. Por lo tanto, los parámetros deben ser estimados de manera consistente para que la aproximación asintótica sea válida. En caso de que los datos provistos por las distintas fuentes de información sean declarados incompatibles, se podría considerar como alternativa el que haya ocurrido un cambio estructural en la serie en el tiempo $N+1$ o que los datos observados al tiempo $N+1$ sean atípicos.

3.5 Pronósticos restringidos para la ESGRM

En el contexto de la ESGRM, la información adicional a la serie histórica serán los datos observados al tiempo $t = N + 1$, y tales datos no están sujetos a incertidumbre, por lo cual la matriz Σ_u siempre será la matriz 0. Los pronósticos irrestrictos obtenidos del modelo VAR, están dados por la esperanza condicional

$$E(\nabla \mathbf{Z}_{N+1}|\mathbf{Z}) = (E(\text{DTNC}_{N+1}|\mathbf{Z}), E(\text{DTPP}_{N+1}|\mathbf{Z}), E(\text{DTPC}_{N+1}|\mathbf{Z}))'. \quad (3.30)$$

Mientras que, respecto a los datos faltantes, se pueden presentar los casos que se enumeran a continuación, en términos de los arreglos \mathbf{C} , \mathbf{Y} y $\nabla \hat{\mathbf{Z}}_{N+1} = (\text{DT}\hat{\text{NC}}_{N+1}, \text{DT}\hat{\text{PP}}_{N+1}, \text{DT}\hat{\text{PC}}_{N+1})'$, que se usan en las expresiones (3.15), (3.16) y (3.29).

Caso 1. Ningún dato es faltante en $t = N + 1$. Entonces $\mathbf{C} = \mathbf{I}_3$, $\mathbf{Y} = \nabla \mathbf{Z}_{N+1}$ y el pronóstico restringido resulta ser $\nabla \hat{\mathbf{Z}}_{N+1} = \mathbf{Y}$, con $E\hat{\mathbf{C}}\mathbf{M}(\nabla \hat{\mathbf{Z}}_{N+1}) = 0$. Además, el estadístico calculado toma la forma

$$K_{\text{calc}} = [\mathbf{Y} - E(\nabla \mathbf{Z}_{N+1} | \mathbf{Z})]' \hat{\Sigma}_a [\mathbf{Y} - E(\nabla \mathbf{Z}_{N+1} | \mathbf{Z})],$$

y con él se puede validar si, en conjunto, los datos observados de las tres variables son congruentes con lo esperado (el pronóstico irrestricto), de acuerdo con el comportamiento histórico que se captura mediante el modelo. Para ello se debe comparar el valor de K_{calc} contra la distribución Ji-cuadrada con 3 grados de libertad.

Caso 2. El único dato faltante es $DTNC_{N+1}$. En este caso, la fórmula del pronóstico restringido se aplica con

$$\mathbf{C} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ y } \mathbf{Y} = \begin{pmatrix} DTPP_{N+1} \\ DTPC_{N+1} \end{pmatrix}$$

y así se obtiene el pronóstico restringido

$$\nabla \mathbf{Z}_{N+1} = \begin{pmatrix} E(DTNC_{N+1} | \mathbf{Z}) + \frac{\sigma_{12}\sigma_3^2 - \sigma_{13}\sigma_{23}}{\sigma_2^2\sigma_3^2 - \sigma_{23}^2} [DTPP_{N+1} - E(DTPP_{N+1} | \mathbf{Z})] \\ + \frac{\sigma_{13}\sigma_2^2 - \sigma_{12}\sigma_{32}}{\sigma_2^2\sigma_3^2 - \sigma_{23}^2} [DTPC_{N+1} - E(DTPC_{N+1} | \mathbf{Z})] \\ DTPP_{N+1} \\ DTPC_{N+1} \end{pmatrix},$$

junto con la estimación de su error cuadrático medio, que es una matriz de dimensión 3×3 con un solo elemento distinto de cero en su entrada (1,1). Dicho elemento es el error cuadrático medio del pronóstico restringido de $DTNC_{N+1}$, y toma el valor

$$E\hat{CM}(DTNC_{N+1}) = \sigma_1^2 - (\sigma_{12}^2\sigma_3^2 - 2\sigma_{12}\sigma_{13}\sigma_{23} + \sigma_{13}^2\sigma_2^2) / (\sigma_2^2\sigma_3^2 - \sigma_{23}^2).$$

Asimismo, el estadístico de compatibilidad está dado por

$$K_{\text{calc}} = \{ \sigma_3^2 [DTPP_{N+1} - E(DTPP_{N+1} | \mathbf{Z})]^2 - 2\sigma_{23} [DTPP_{N+1} - E(DTPP_{N+1} | \mathbf{Z})] \\ \times [DTPC_{N+1} - E(DTPC_{N+1} | \mathbf{Z})] + \sigma_2^2 [DTPC_{N+1} - E(DTPC_{N+1} | \mathbf{Z})]^2 \} / (\sigma_2^2\sigma_3^2 - \sigma_{23}^2),$$

el cual permite validar que los datos observados de las dos variables no discrepen significativamente de los pronósticos irrestrictos producidos por el modelo. Ahora K_{calc} se compara contra la distribución Ji-cuadrada con 2 grados de libertad.

Caso 3. El dato faltante es $DTPP_{N+1}$, de tal manera que

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} DTNC_{N+1} \\ DTPC_{N+1} \end{pmatrix} \text{ y}$$

$$\nabla \mathbf{Z}_{N+1} = \begin{pmatrix} \text{DTNC}_{N+1} \\ E(\text{DTPP}_{N+1} | \mathbf{Z}) + \frac{\sigma_{12}\sigma_3^2 - \sigma_{23}\sigma_{13}}{\sigma_1^2\sigma_3^2 - \sigma_{13}^2} [\text{DTNC}_{N+1} - E(\text{DTNC}_{N+1} | \mathbf{Z})] \\ + \frac{\sigma_{23}\sigma_1^2 - \sigma_{12}\sigma_{13}}{\sigma_1^2\sigma_3^2 - \sigma_{13}^2} [\text{DTPC}_{N+1} - E(\text{DTPC}_{N+1} | \mathbf{Z})] \\ \text{DTPC}_{N+1} \end{pmatrix}.$$

Su matriz de error cuadrático medio contiene sólo un elemento distinto de cero, el (2,2)

$$E\hat{\text{CM}}(\text{DT}\hat{\text{P}}_{N+1}) = \sigma_2^2 - (\sigma_{23}^2\sigma_1^2 - 2\sigma_{12}\sigma_{13}\sigma_{23} + \sigma_{12}^2\sigma_3^2)/(\sigma_1^2\sigma_3^2 - \sigma_{13}^2).$$

Por su lado, el estadístico de compatibilidad es

$$K_{\text{calc}} = \{\sigma_3^2[\text{DTNC}_{N+1} - E(\text{DTNC}_{N+1} | \mathbf{Z})]^2 - 2\sigma_{13}[\text{DTNC}_{N+1} - E(\text{DTNC}_{N+1} | \mathbf{Z})] \\ \times [\text{DTPC}_{N+1} - E(\text{DTPC}_{N+1} | \mathbf{Z})] + \sigma_1^2[\text{DTPC}_{N+1} - E(\text{DTPC}_{N+1} | \mathbf{Z})]^2\}/(\sigma_1^2\sigma_3^2 - \sigma_{13}^2).$$

Caso 4. Falta el dato de DTPC_{N+1} , entonces

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} \text{DTNC}_{N+1} \\ \text{DTPP}_{N+1} \end{pmatrix} \text{ y}$$

$$\nabla \mathbf{Z}_{N+1} = \begin{pmatrix} \text{DTNC}_{N+1} \\ \text{DTPP}_{N+1} \\ E(\text{DTPC}_{N+1} | \mathbf{Z}) + \frac{\sigma_{13}\sigma_2^2 - \sigma_{23}\sigma_{12}}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2} [\text{DTNC}_{N+1} - E(\text{DTNC}_{N+1} | \mathbf{Z})] \\ + \frac{\sigma_{23}\sigma_1^2 - \sigma_{13}\sigma_{21}}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2} [\text{DTPP}_{N+1} - E(\text{DTPP}_{N+1} | \mathbf{Z})] \end{pmatrix}.$$

De nuevo, la matriz de error cuadrático medio de este vector de pronósticos restringidos sólo contiene un elemento distinto de cero, el (3,3) en este caso. Este elemento es

$$E\hat{\text{CM}}(\text{DT}\hat{\text{P}}_{N+1}) = \sigma_3^2 - (\sigma_{23}^2\sigma_1^2 - 2\sigma_{12}\sigma_{13}\sigma_{23} + \sigma_{13}^2\sigma_2^2)/(\sigma_1^2\sigma_2^2 - \sigma_{12}^2),$$

mientras que el estadístico de compatibilidad es de la forma

$$K_{\text{calc}} = \{\sigma_2^2[\text{DTNC}_{N+1} - E(\text{DTNC}_{N+1} | \mathbf{Z})]^2 - 2\sigma_{12}[\text{DTNC}_{N+1} - E(\text{DTPC}_{N+1} | \mathbf{Z})] \\ \times [\text{DTPP}_{N+1} - E(\text{DTPP}_{N+1} | \mathbf{Z})] + \sigma_1^2[\text{DTPP}_{N+1} - E(\text{DTPP}_{N+1} | \mathbf{Z})]^2\}/(\sigma_1^2\sigma_2^2 - \sigma_{12}^2).$$

Caso 5. Los datos faltantes son DTNC_{N+1} y DTPP_{N+1} , así que $\mathbf{C} = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$, $\mathbf{Y} = \text{DTPC}_{N+1}$ y

$$\nabla \hat{\mathbf{Z}}_{N+1} = \begin{pmatrix} E(DTNC_{N+1} | \mathbf{Z}) + \frac{\sigma_{13}}{\sigma_3^2} [DTPC_{N+1} - E(DTPC_{N+1} | \mathbf{Z})] \\ E(DTPP_{N+1} | \mathbf{Z}) + \frac{\sigma_{23}}{\sigma_3^2} [DTPC_{N+1} - E(DTPC_{N+1} | \mathbf{Z})] \\ DTPC_{N+1} \end{pmatrix}.$$

Ahora, la matriz estimada de error cuadrático medio del vector de pronósticos contiene un bloque de dimensión 2×2 , de elementos distintos de cero. Dicho bloque está dado por

$$E\hat{CM} \begin{pmatrix} DT\hat{N}C_{N+1} \\ DT\hat{P}P_{N+1} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 - \sigma_{13}^2 / \sigma_3^2 & \sigma_{12} - \sigma_{13}\sigma_{23} / \sigma_3^2 \\ \sigma_{12} - \sigma_{13}\sigma_{23} / \sigma_3^2 & \sigma_2^2 - \sigma_{23}^2 / \sigma_3^2 \end{pmatrix}$$

y el correspondiente estadístico de compatibilidad resulta ser

$$K_{\text{calc}} = [DTPC_{N+1} - E(DTPC_{N+1} | \mathbf{Z})]^2 / \sigma_3^2,$$

con el cual se valida que no haya discrepancia significativa entre el dato observado y su pronóstico irrestricto, al comparar el valor de K_{calc} contra la distribución Ji-cuadrada con 1 grado de libertad.

Caso 6. Faltan los datos de $DTNC_{N+1}$ y $DTPC_{N+1}$, por lo cual $C = (0 \ 1 \ 0)$, $\mathbf{Y} = DTPP_{N+1}$,

$$\nabla \hat{\mathbf{Z}}_{N+1} = \begin{pmatrix} E(DTNC_{N+1} | \mathbf{Z}) + \frac{\sigma_{12}}{\sigma_2^2} [DTPP_{N+1} - E(DTPP_{N+1} | \mathbf{Z})] \\ DTPP_{N+1} \\ E(DTPC_{N+1} | \mathbf{Z}) + \frac{\sigma_{23}}{\sigma_2^2} [DTPP_{N+1} - E(DTPP_{N+1} | \mathbf{Z})] \end{pmatrix}$$

y

$$E\hat{CM} \begin{pmatrix} DT\hat{N}C_{N+1} \\ DT\hat{P}C_{N+1} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 - \sigma_{12}^2 / \sigma_2^2 & \sigma_{13} - \sigma_{12}\sigma_{23} / \sigma_2^2 \\ \sigma_{13} - \sigma_{12}\sigma_{23} / \sigma_2^2 & \sigma_3^2 - \sigma_{23}^2 / \sigma_2^2 \end{pmatrix}.$$

El correspondiente estadístico de compatibilidad es entonces

$$K_{\text{calc}} = [DTPP_{N+1} - E(DTPP_{N+1} | \mathbf{Z})]^2 / \sigma_2^2.$$

Caso 7. Los datos que faltan son $DTPP_{N+1}$ y $DTPC_{N+1}$, entonces $C = (1 \ 0 \ 0)$, $\mathbf{Y} = DTNC_{N+1}$,

$$\nabla \hat{\mathbf{Z}}_{N+1} = \begin{pmatrix} \text{DTNC}_{N+1} \\ E(\text{DTPP}_{N+1} | \mathbf{Z}) + \frac{\hat{\sigma}_{12}}{\hat{\sigma}_1^2} [\text{DTNC}_{N+1} - E(\text{DTPP}_{N+1} | \mathbf{Z})] \\ E(\text{DTPC}_{N+1} | \mathbf{Z}) + \frac{\hat{\sigma}_{13}}{\hat{\sigma}_1^2} [\text{DTNC}_{N+1} - E(\text{DTPC}_{N+1} | \mathbf{Z})] \end{pmatrix}$$

y

$$E\hat{\text{CM}} \begin{pmatrix} \text{DTPP}_{N+1} \\ \text{DTPC}_{N+1} \end{pmatrix} = \begin{pmatrix} \hat{\sigma}_2^2 - \hat{\sigma}_{12}^2 / \hat{\sigma}_1^2 & \hat{\sigma}_{23} - \hat{\sigma}_{12}\hat{\sigma}_{13} / \hat{\sigma}_1^2 \\ \hat{\sigma}_{23} - \hat{\sigma}_{12}\hat{\sigma}_{13} / \hat{\sigma}_1^2 & \hat{\sigma}_3^2 - \hat{\sigma}_{13}^2 / \hat{\sigma}_1^2 \end{pmatrix},$$

junto con su correspondiente estadístico de compatibilidad

$$K_{\text{calc}} = [\text{DTNC}_{N+1} - E(\text{DTNC}_{N+1} | \mathbf{Z})]^2 / \hat{\sigma}_1^2.$$

Caso 8. Todos los datos son faltantes. En este caso no hay restricción por satisfacer y, por lo tanto, el mejor pronóstico es el irrestricto, $\nabla \hat{\mathbf{Z}}_{N+1} = E(\nabla \mathbf{Z}_{N+1} | \mathbf{Z})$, cuyo error cuadrático medio estimado está dado por $\hat{\Sigma}_a$, o sea

$$\nabla \hat{\mathbf{Z}}_{N+1} = \begin{pmatrix} E(\text{DTNC}_{N+1} | \mathbf{Z}) \\ E(\text{DTPP}_{N+1} | \mathbf{Z}) \\ E(\text{DTPC}_{N+1} | \mathbf{Z}) \end{pmatrix}, \text{ con } E\hat{\text{CM}} \begin{pmatrix} \text{DTNC}_{N+1} \\ \text{DTPP}_{N+1} \\ \text{DTPC}_{N+1} \end{pmatrix} = \begin{pmatrix} \hat{\sigma}_1^2 & \hat{\sigma}_{12} & \hat{\sigma}_{13} \\ \hat{\sigma}_{12} & \hat{\sigma}_2^2 & \hat{\sigma}_{23} \\ \hat{\sigma}_{13} & \hat{\sigma}_{23} & \hat{\sigma}_3^2 \end{pmatrix}.$$

El estadístico de compatibilidad no es aplicable en este caso.

Lo que debe notarse en las expresiones previas es, en primer lugar, que los pronósticos restringidos para los valores observados toman exactamente dichos valores, de manera que las restricciones impuestas se satisfacen exactamente. En segundo término, los errores cuadráticos medios estimados de los pronósticos restringidos son menores que los correspondientes para los pronósticos irrestrictos. Además, las expresiones mostradas permiten apreciar que en los cálculos se hace uso de toda la información disponible para estimar los datos faltantes y que el estadístico de compatibilidad pondera con el recíproco de las varianzas y de las covarianzas a las diferencias entre los valores observados y los pronosticados de manera irrestricta, todo lo cual produce resultados muy razonables.

4. CONSTRUCCIÓN DEL MODELO VAR

La construcción del modelo se realizó de acuerdo con el comportamiento observado en los datos y a partir de los métodos estándar para este tipo de situaciones. Es decir,

primero se decidió el grado de diferenciación apropiado para volver estacionarias las series. Para ello se utilizaron las gráficas de las series transformadas y se aplicó el método de diferencia de la variable (véase al respecto el texto de Guerrero, 2003) de manera que la desviación estándar mínima de cada una de las series y de sus respectivas diferencias, indica el grado de diferenciación que debe aplicarse. Desde luego, pudo haberse empleado un método econométrico más formal para determinar si existen raíces unitarias en la serie transformada (lo que equivale a decir que se requiere aplicar diferencias para volver estacionaria la serie respectiva), sin embargo, usar ese tipo de pruebas se vuelve demasiado casuístico y dificulta la aplicación masiva del método a todos los rastros de la República Mexicana.

4.1 Determinación del grado de diferenciación

Los resultados del método de diferencia de la variable se muestran en los cuadros 4.1 a 4.3 a continuación, para algunos rastros de Aguascalientes. En ellos se puede observar que, en general, la serie original no es estacionaria, mientras que la primera diferencia produce las desviaciones estándar mínimas en la gran mayoría de los casos. Esto indica que la estacionariedad se logra con la primera diferencia. Además, cuando el mínimo no se presenta en la primera diferencia, ocurre en la serie original y esto sólo para alguna de las variables de alguna especie y rastro. Es claro que se podría sobrediferenciar la serie respectiva, pero igualmente se mantendría la estacionariedad al aplicar la diferencia. Así, se está ya en posibilidad de escoger el número de retrasos de los modelos VAR para las variables, por especie y por rastro.

Cuadro 4.1 Desviación estándar de acuerdo al grado de diferenciación (Rastro X3)

Variable	Especie	Número de diferencias		
		0	1	2
NC	Bovinos	45.23	40.29	56.80
	Porcinos	42.74	25.70	38.57
PP	Bovinos	21119.94	18537.69	27322.43
	Porcinos	6209.75	3708.76	5545.01
PC	Bovinos	13575.32	9593.04	14327.54
	Porcinos	2945.44	2720.19	4438.84

Cuadro 4.2 Desviación estándar de acuerdo al grado de diferenciación (Rastro X7)

Variable	Especie	Número de diferencias		
		0	1	2
NC	Bovinos	34.01882	26.70051	40.15907
	Porcinos	59.28407	38.82589	59.31340
	Ovinos	15.00514	10.52564	16.54329
	Caprinos	10.82237	10.41345	16.73383
PP	Bovinos	19500.28	15395.25	25283.03
	Porcinos	6706.947	4470.787	7000.223
	Ovinos	725.4872	468.4863	777.0064
	Caprinos	320.6480	366.1169	558.3796
PC	Bovinos	8122.266	6696.902	11016.37
	Porcinos	4528.217	3287.474	5246.181
	Ovinos	373.4184	235.4393	387.3680
	Caprinos	137.4493	158.9118	256.3080

Cuadro 4.3 Desviación estándar de acuerdo al grado de diferenciación (Rastro X8)

Variable	Especie	Número de diferencias		
		0	1	2
NC	Bovinos	422.5391	333.8587	532.8909
	Porcinos	1315.598	802.3517	1237.642
	Ovinos	121.3489	128.0909	197.4474
	Caprinos	206.0397	205.2253	313.8000
PP	Bovinos	167187.3	136136.0	224561.0
	Porcinos	142422.0	88340.29	137146.9
	Ovinos	5885.563	4819.495	7323.128
	Caprinos	7272.525	6984.003	10706.32
PC	Bovinos	100253.4	65870.83	105970.9
	Porcinos	92477.00	58963.06	91627.71
	Ovinos	3396.992	2834.050	4243.361
	Caprinos	3974.419	3610.400	5577.390

4.2 Pruebas de razón de verosimilitud

El orden del modelo VAR para cada especie, en cada uno de los rastros, se determinó mediante pruebas de razón de verosimilitud. Estas pruebas permiten probar las hipótesis sucesivas

$$H_0: \text{el orden es } p \text{ vs. } H_A: \text{el orden es } p-1$$

con un valor inicial elegido como $p = 6$. Ejemplos de los resultados de la aplicación de dichas pruebas se muestran en los cuadros 4.1 al 4.3, en donde se aprecia que, en términos generales, la especificación genérica del modelo debe ser de orden $p = 3$, ya que es para ese orden del VAR para el cual comúnmente se obtuvo el primer valor significativo de la estadística χ^2 , al ir reduciendo el número de retrasos del modelo. Adicionalmente, debe apreciarse que en todos los casos se usó la misma especificación del modelo en lo que toca a los elementos determinísticos. La forma usada es

$$\nabla \mathbf{Z}_t = \Lambda \mathbf{D}_t + \Pi_1 \nabla \mathbf{Z}_{t-1} + \dots + \Pi_p \nabla \mathbf{Z}_{t-p} + \mathbf{a}_t$$

en donde el vector \mathbf{D}_t tiene como elementos solamente a las variables artificiales para capturar los efectos estacionales de cada serie.

Cuadro 4.4 Pruebas de razón de verosimilitud (Rastro X3)

p (retrasos)	Bovino		Porcino	
	χ^2	Signif.	χ^2	Signif.
6	22.38	0.01	19.28	0.02
5	13.28	0.15	11.18	0.26
4	4.99	0.84	4.94	0.84
3	19.48	0.02	21.39	0.01
2	37.36	0.00	21.94	0.01

Cuadro 4.5 Pruebas de razón de verosimilitud (Rastro X5)

p (retrasos)	Bovino		Porcino		Ovino	
	χ^2	Signif.	χ^2	Signif.	χ^2	Signif.
6	3.79	0.92	3.94	0.92	11.11	0.27
5	4.57	0.87	5.39	0.80	9.60	0.38
4	3.57	0.94	8.61	0.47	9.14	0.42
3	37.48	0.00	8.02	0.53	25.21	0.002
2	14.52	0.11	29.02	0.00	19.80	0.02

Cuadro 4.6 Pruebas de razón de verosimilitud (Rastro X8)

p (retrasos)	Bovino		Porcino		Ovino		Caprino	
	χ^2	Signif.	χ^2	Signif.	χ^2	Signif.	χ^2	Signif.
6	8.09	0.53	8.21	0.51	7.75	0.56	11.93	0.22
5	10.85	0.29	10.01	0.35	14.37	0.11	9.02	0.44
4	9.28	0.41	16.02	0.07	13.70	0.13	16.77	0.05
3	11.53	0.24	14.27	0.11	17.64	0.04	20.87	0.01
2	30.21	0.00	30.06	0.00	20.64	0.01	24.85	0.00

4.3 Estimación del modelo

Una vez determinado el orden del VAR, se procedió a estimar el modelo. Las tres ecuaciones que forman el modelo son

$$\begin{aligned} \text{DTNC}_t = & \pi_{1,11}\text{DTNC}_{t-1} + \dots + \pi_{3,11}\text{DTNC}_{t-3} + \pi_{1,21}\text{DTPP}_{t-1} + \dots + \pi_{3,21}\text{DTPP}_{t-3} \\ & + \pi_{1,31}\text{DTPC}_{t-1} + \dots + \pi_{3,31}\text{DTPC}_{t-3} + \sum_{i=2}^{12} \lambda_i D_{i,t} + a_{\text{DTNC},t} \end{aligned}$$

$$\begin{aligned} \text{DTPP}_t = & \pi_{1,12}\text{DTNC}_{t-1} + \dots + \pi_{3,12}\text{DTNC}_{t-3} + \pi_{1,22}\text{DTPP}_{t-1} + \dots + \pi_{3,22}\text{DTPP}_{t-3} \\ & + \pi_{1,32}\text{DTPC}_{t-1} + \dots + \pi_{3,32}\text{DTPC}_{t-3} + \sum_{i=2}^{12} \lambda_i D_{i,t} + a_{\text{DTPP},t} \end{aligned}$$

$$\begin{aligned} \text{DTPC}_t = & \pi_{1,13}\text{DTNC}_{t-1} + \dots + \pi_{3,13}\text{DTNC}_{t-3} + \pi_{1,23}\text{DTPP}_{t-1} + \dots + \pi_{3,23}\text{DTPP}_{t-3} \\ & + \pi_{1,33}\text{DTPC}_{t-1} + \dots + \pi_{3,33}\text{DTPC}_{t-3} + \sum_{i=2}^{12} \lambda_i D_{i,t} + a_{\text{DTPC},t} \end{aligned}$$

con $t = 6, \dots, N$, donde la estacionalidad se captura mediante variables artificiales centradas (en el sentido de que la suma de los valores de $D_{i,t}$ dentro de cada año es cero, para $i = 2, \dots, 12$), es decir,

$$\begin{aligned} D_{2,t} = \begin{cases} 11/12 & \text{si } t = \text{febrero} \\ -1/12 & \text{e.o.c.} \end{cases}, \dots, D_{11,t} = \begin{cases} 11/12 & \text{si } t = \text{noviembre} \\ -1/12 & \text{e.o.c.} \end{cases} \quad y \\ D_{12,t} = \begin{cases} 11/12 & \text{si } t = \text{diciembre} \\ -1/12 & \text{e.o.c.} \end{cases} \end{aligned}$$

La estimación de las ecuaciones del modelo se puede llevar a cabo de manera eficiente mediante el método de Mínimos Cuadrado Ordinarios aplicado a cada una de las ecuaciones por separado, lo cual hace que este proceso sea bastante simple de realizar. Cabe mencionar que la parte computacional inicial, relacionada con el análisis de datos y la construcción de los modelos, fue realizada con el apoyo del paquete de

cómputo RATS, versión 5 (<http://www.estima.com>). Los resultados de la estimación de los modelos VAR por rastro y especie, produjeron resultados del tipo mostrado en los cuadros 4.7 a 4.10 para algunos rastros seleccionados, con los cuales se ejemplifican muy bien los resultados obtenidos en general para todos los rastros. En estos cuadros se aprecia que la estacionalidad tiene efectos significativos para explicar el comportamiento de las tres variables en consideración, sin importar la especie. Por su lado, los valores de los coeficientes de determinación ajustados por grados de libertad son relativamente altos para la variable DTNC (alrededor de 65%), y sensiblemente más bajos (alrededor de 35%) para DTPP y DTPC. La explicación de las variables en buena medida se debe a su propia inercia, como lo muestran los valores significativos de los estadísticos F, los cuales se usan para probar la hipótesis nula de que la correspondiente variable explicativa no tiene efectos significativos sobre el comportamiento de la variable que se desea explicar.

Cuadro 4.7 Resultados de la estimación del modelo VAR (3): Especie Bovinos

Variables por explicar	Variables explicativas (estadística F)				\overline{R}^2
	DTNC	DTPP	DTPC	Estac.	
Rastro X1					
DTNC	25.29***	0.35	0.58	*	0.57
DTPP	0.12	8.52***	1.79	**	0.36
DTPC	0.66	3.97***	1.01	*	0.39
Rastro X5					
DTNC	30.59***	0.30	0.45	***	0.70
DTPP	1.77	3.20**	0.05	**	0.34
DTPC	1.50	0.40	1.37	*	0.29
Rastro X6					
DTNC	13.85***	6.09***	5.69***	***	0.68
DTPP	1.94	2.06	2.26*	***	0.35
DTPC	1.43	0.73	7.32***	--	0.31

(***) Significativo al 1%, (**) Significativo al 5%, (*) Significativo al 10%, (--) No significativo al 10%

Cuadro 4.8 Resultados de la estimación del modelo VAR (3): Especie Porcinos

Variables por explicar	Variables explicativas (estadística F)				\overline{R}^2
	DTNC	DTPP	DTPC	Estac.	
Rastro X3					
DTNC	18.59***	0.98	0.90	***	0.74
DTPP	0.24	3.55**	0.43	--	0.15
DTPC	0.20	0.88	3.82**	**	0.21
Rastro X4					
DTNC	23.95***	5.45***	3.63**	***	0.75
DTPP	1.64	2.54*	0.18	--	0.32
DTPC	0.37	0.19	3.59**	--	0.19
Rastro X8					
DTNC	10.27***	0.66	0.52	***	0.82
DTPP	1.49	5.27***	0.54	--	0.26
DTPC	1.96	0.88	8.25***	--	0.37

(***) Significativo al 1%, (**) Significativo al 5%, (*) Significativo al 10%, (--) No significativo al 10%

Cuadro 4.9 Resultados de la estimación del modelo VAR (3): Especie Ovinos

Variables por explicar	Variables explicativas (estadística F)				\overline{R}^2
	DTNC	DTPP	DTPC	Estac.	
Rastro X2					
DTNC	13.61***	1.13	1.83	***	0.65
DTPP	0.04	0.88	1.37	**	0.19
DTPC	0.18	1.42	2.71**	*	0.22
Rastro X5					
DTNC	3.88***	4.05**	5.63***	*	0.49
DTPP	0.15	2.32*	0.87	*	0.08
DTPC	1.08	0.25	0.37	*	0.12
Rastro X6					
DTNC	17.22***	0.97	1.64	***	0.56
DTPP	1.04	2.21*	0.17	***	0.39
DTPC	2.89**	0.93	6.01***	**	0.38

(***) Significativo al 1%, (**) Significativo al 5%, (*) Significativo al 10%, (--) No significativo al 10%

Cuadro 4.10 Resultados de la estimación del modelo VAR (3): Especie Caprinos

Variables por explicar	Variables explicativas (estadística F)				\bar{R}^2
	DTNC	DTPP	DTPC	Estac.	
Rastro X6					
DTNC	13.24***	3.65**	2.08	--	0.51
DTPP	1.92	4.14***	2.80**	--	0.39
DTPC	1.77	2.25*	4.41***	--	0.45
Rastro X8					
DTNC	19.18***	0.62	3.42**	***	0.81
DTPP	1.20	8.93***	2.51*	***	0.40
DTPC	1.14	0.63	2.96**	***	0.25

(***) Significativo al 1%, (**) Significativo al 5%, (*) Significativo al 10%, (--) No significativo al 10%

5. APLICACIÓN DEL METODO DE EDICIÓN E IMPUTACIÓN

En esta sección se muestran algunos de los resultados de la aplicación del método propuesto en la práctica. Con esto se pretende simplemente ilustrar el tipo de resultados que se obtienen al realizar edición e imputación de datos.

5.1 Edición de valores incongruentes

En el Cuadro 5.1 se tiene un ejemplo en el cual el rastro no tuvo datos faltantes para ninguna de las dos especies de ganado que ahí se sacrifica (bovino y porcino), en ninguno de los meses considerados. Debe notarse, sin embargo, que los datos observados para el mes de enero resultaron ser notoriamente más altos que los esperados según el comportamiento histórico de las variables capturado por el modelo, para ambas especies. Esto se aprecia en el hecho de que el estadístico de compatibilidad tomó el valor $K_{\text{calc}} = 6.94$ en el caso de bovinos, y $K_{\text{calc}} = 14.37$ en el caso de porcinos, los cuales resultan ser significativos (al nivel de significancia del 5%) al compararse con una distribución Ji-cuadrada con 3 grados de libertad. Para bovinos volvieron a ocurrir discrepancias significativas en los meses de noviembre y diciembre, y para porcinos en abril y diciembre. Estas discrepancias se aprecia que son causadas fundamentalmente por cambios en la tendencia de las variables. Desde luego, los pronósticos restringidos tienen en cuenta los valores observados y los corrigen cuando esto es necesario, de acuerdo con la edición automática que realiza el procedimiento. De manera que los datos de este rastro permiten ilustrar básicamente la parte de edición del procedimiento.

Cuadro 5.1 Resultados del procedimiento en tiempo real para el rastro X2 (Año 2004)

Mes	Datos originales			Estad. de	Pronósticos irrestrictos			Pronósticos restringidos		
	NC	PP	PC	Comp. (sig.)	NC	PP	PC	NC	PP	PC
BOVINOS										
Ene	25	10000	4500	6.94 (0.03)	7	2153	1234	25	10000	5190
Feb	11	4400	1466	3.56 (0.17)	26	6659	3509	11	4400	2200
Mar	12	4800	1600	2.02 (0.36)	26	9500	4602	12	4800	2327
Abr	22	8800	2933	0.25 (0.88)	24	11320	6352	22	8800	5328
May	14	5600	1867	2.21 (0.33)	19	10263	4901	14	5600	2453
Jun	13	5200	1733	0.26 (0.88)	16	7090	4213	13	5200	3228
Jul	12	4800	1600	0.52 (0.77)	18	6619	3736	12	4800	2688
Ago	13	5200	1733	0.81 (0.67)	8	3023	1665	13	5200	2744
Sep	8	3200	1066	0.46 (0.80)	9	4549	2264	8	3200	1647
Oct	8	3200	1066	0.04 (0.98)	9	3628	1416	8	3200	1335
Nov	3	270	135	15.68 (0.00)	7	2716	1645	3	750	381
Dic	4	1000	500	6.39 (0.04)	2	471	264	4	1000	503
PORCINOS										
Ene	124	11160	6138	14.37 (0.00)	37	3305	2068	124	11160	7280
Feb	51	4560	1530	1.50 (0.47)	59	5273	3396	51	4590	3046
Mar	60	5400	1800	2.03 (0.36)	61	5527	3955	60	5400	3694
Abr	41	3690	1230	7.35 (0.03)	103	9260	5833	41	3690	2362
May	66	5940	1980	0.18 (0.91)	64	5799	3476	66	5940	3765
Jun	59	5310	1770	2.26 (0.32)	49	4384	2829	59	5310	3576
Jul	42	3780	1260	4.23 (0.12)	79	7227	5292	42	3780	2641
Ago	54	4860	1620	1.17 (0.56)	38	3405	2287	54	4860	3224
Sep	52	4680	3276	0.41 (0.81)	42	3743	2469	52	4680	3120
Oct	44	3960	1320	0.98 (0.61)	44	3972	2780	44	3960	2683
Nov	35	3150	1575	0.15 (0.93)	37	3295	2232	35	3150	2109
Dic	40	3920	1960	6.83 (0.03)	34	3031	1993	40	3920	2604

5.2 Edición e imputación de valores faltantes

En los cuadros 5.2 y 5.3 se muestran ejemplos en donde se efectúa la imputación de datos, además de la edición de los mismos, cuando ésta es requerida. Respecto al Cuadro 5.2, el dato faltante para las cuatro especies y todos los meses fue el de PC. Después de contactar al rastro para saber el motivo de esta falta repetida de la información, la respuesta fue que la báscula para pesar la carne después del sacrificio se había descompuesto y por ello no se podía proporcionar el dato. Es de notar que en este rastro hay menos datos discrepantes por especie que en el caso del Cuadro 5.1, de hecho, sólo aparece como incompatible la pareja de datos de porcinos en diciembre, las de ovinos en enero, marzo y septiembre, y la pareja de observaciones de caprinos en septiembre.

Cuadro 5.2 Resultados del procedimiento en tiempo real para el rastro X6 (Año 2004)

Mes	Datos originales			Estad. de comp. (sig.)	Pronósticos irrestrictos			Pronósticos restringidos		
	NC	PP	PC		NC	PP	PC	NC	PP	PC
BOVINOS										
Ene	139	76450	ND	4.80 (0.09)	186	101955	56414	139	76450	42049
Feb	104	57200	ND	1.24 (0.54)	88	48635	26746	104	57200	31461
Mar	107	58850	ND	0.00 (1.00)	106	58434	32139	107	58850	32367
Abr	112	61600	ND	0.77 (0.68)	98	53823	29599	112	61600	33878
May	134	73700	ND	1.07 (0.59)	117	64253	35334	134	73700	40533
Jun	127	69850	ND	0.46 (0.80)	137	75350	41446	127	69850	38409
Jul	130	71500	ND	0.14 (0.93)	126	69191	38050	130	71500	39322
Ago	115	63250	ND	1.37 (0.50)	137	75379	41451	115	63250	34780
Sep	109	59950	ND	0.16 (0.92)	104	57218	31469	109	59950	32969
Oct	95	52250	ND	4.56 (0.10)	123	67821	37293	95	52250	28729
Nov	106	58300	ND	1.27 (0.53)	90	49386	27160	106	58300	32061
Dic	120	66000	ND	0.93 (0.63)	120	65781	36215	120	66000	36293
PORCINOS										
Ene	185	20350	ND	0.90 (0.64)	223	23984	17167	185	20350	14343
Feb	133	14630	ND	0.05 (0.98)	129	14256	9926	133	14630	10222
Mar	126	13860	ND	2.33 (0.31)	123	14024	9672	126	13860	9619
Abr	138	15180	ND	0.11 (0.95)	132	14312	10036	138	15180	10597
May	174	19140	ND	1.24 (0.54)	142	15133	10378	174	19140	13093
Jun	156	17160	ND	0.04 (0.98)	150	16599	11400	156	17160	11807
Jul	166	18260	ND	0.07 (0.97)	174	19241	13469	166	18260	12804
Ago	143	15730	ND	0.47 (0.79)	165	18208	12672	143	15730	10939
Sep	169	18590	ND	1.72 (0.42)	133	14155	9989	169	18590	12980
Oct	168	18480	ND	1.02 (0.60)	205	22827	15859	168	18480	12857
Nov	162	17820	ND	0.69 (0.71)	144	15300	10773	162	17820	12405
Dic	209	22990	ND	10.03 (0.01)	214	23792	16449	209	22990	16705
OVINOS										
Ene	50	2000	ND	12.45 (0.00)	86	3872	2059	50	2000	907
Feb	37	1480	ND	2.89 (0.24)	48	1609	825	37	1480	720
Mar	45	2025	ND	6.08 (0.05)	28	1228	572	45	2025	1067
Abr	55	2475	ND	2.43 (0.30)	39	1740	780	55	2475	1225
May	56	2520	ND	2.74 (0.25)	61	2658	1319	56	2520	1342
Jun	57	2280	ND	4.17 (0.12)	48	2164	1168	57	2280	1243
Jul	59	2655	ND	2.49 (0.29)	71	3143	1614	59	2655	1423
Ago	50	2250	ND	0.50 (0.78)	57	2577	1329	50	2250	1179
Sep	53	2120	ND	7.22 (0.03)	42	1908	1011	53	2120	1022
Oct	52	2340	ND	1.91 (0.38)	53	2355	1197	52	2340	1253
Nov	49	2205	ND	0.41 (0.81)	44	1954	1033	49	2205	1182
Dic	71	2840	ND	4.06 (0.13)	62	2783	1428	71	2840	1330
CAPRINOS										
Ene	21	735	ND	0.19 (0.91)	24	804	405	21	735	372
Feb	21	735	ND	0.80 (0.67)	24	970	534	21	735	398
Mar	16	560	ND	0.61 (0.74)	24	823	410	16	560	279
Abr	8	280	ND	1.31 (0.52)	16	562	309	8	280	153
May	15	525	ND	1.94 (0.38)	7	229	107	15	525	246
Jun	9	315	ND	1.52 (0.47)	15	487	247	9	315	161
Jul	12	420	ND	1.23 (0.54)	8	325	171	12	420	216
Ago	12	420	ND	0.53 (0.77)	12	390	188	12	420	198
Sep	14	420	ND	8.26 (0.02)	17	590	334	14	420	212
Oct	18	630	ND	1.68 (0.43)	9	308	182	18	630	373
Nov	15	525	ND	0.21 (0.90)	17	552	301	15	525	291
Dic	5	175	ND	1.13 (0.57)	8	269	140	5	175	92

Los valores imputados de la variable PC, en todos los casos del Cuadro 5.2, tiende a satisfacer aproximadamente los criterios de validación establecidos por las constantes K2 y K4. En los casos en que esto no se cumple, se debe a que en realidad la restricción impuesta es que PC/NC se encuentre dentro del intervalo $[K1 \cdot K3, K2 \cdot K4]$. No obstante, conviene mencionar que los pronósticos restringidos se someten al proceso de edición cuando se reestima el modelo para obtener los pronósticos del mes siguiente. También es de notar que los pronósticos restringidos de la variable PC difieren de los pronósticos irrestrictos para esa misma variable, de manera que no es suficiente con tener un modelo de pronóstico para cada una de las variables por separado, sino que deben considerarse también las interrelaciones que se presenten entre las variables del sistema.

Por su lado, el Cuadro 5.3 muestra otra situación que se presenta con menos frecuencia, pero que ocurrió el año 2004 en uno de los rastros en estudio, esto es, que haya datos faltantes para dos de las tres variables del sistema. Este caso se presentó debido a que el informante se negó a brindar los datos que se le solicitan en el cuestionario, aun después de haber entrado en contacto telefónico con él. Aquí hay necesidad entonces de imputar los datos de dos de las tres variables, a saber, PP y PC, para la especie de ganado porcino y todos los meses del año (los datos faltantes se denotan como ND = no disponible).

Cuadro 5.3 Resultados del procedimiento en tiempo real para el rastro X7 (Año 2004)

Mes	Datos originales			Estad. de comp. (sig.)	Pronósticos irrestrictos			Pronósticos restringidos		
	NC	PP	PC		NC	PP	PC	NC	PP	PC
PORCINOS										
Ene	42	ND	ND	0.07 (0.80)	34	3583	2512	42	4419	3109
Feb	46	ND	ND	0.28 (0.60)	31	3277	2165	46	4990	3352
Mar	56	ND	ND	0.79 (0.38)	29	3009	1970	56	6097	4106
Abr	53	ND	ND	0.19 (0.66)	74	7959	5416	53	5621	3782
May	61	ND	ND	0.04 (0.85)	53	5665	3763	61	6572	4391
Jun	47	ND	ND	0.32 (0.57)	31	3306	2003	47	5150	3207
Jul	48	ND	ND	0.47 (0.49)	80	8700	5866	48	5111	3376
Ago	57	ND	ND	0.06 (0.81)	68	7438	4997	57	6189	4129
Sep	78	ND	ND	0.17 (0.68)	58	6259	4250	78	8566	5865
Oct	41	ND	ND	0.40 (0.53)	65	6929	4657	41	4255	2805
Nov	42	ND	ND	0.18 (0.67)	57	6169	4145	42	4447	2951
Dic	65	ND	ND	0.08 (0.78)	53	5778	3726	65	7159	4670

En este cuadro, se ve que los datos de la variable NC reportados por el informante no resultaron ser incongruentes con la historia en ninguno de los meses y el

procedimiento permitió imputar los datos faltantes de las variables requeridas sin dificultad. Desde luego, hay que reconocer que las cifras obtenidas mediante la imputación, son pronósticos y, por ende, están sujetos a incertidumbre. Esta incertidumbre puede medirse a través del ECM estimado del pronóstico restringido. Sin embargo, se prefirió no presentar estas medidas de variabilidad ya que corresponden a rastros y especies individuales, mientras que los datos que aparecerán en la publicación oficial de la ESGRM son cifras agregadas por especie, de todos los rastros de la entidad federativa.

5.3 Simulaciones

La idea de realizar un pequeño estudio de simulación surgió del deseo de poner a prueba la eficacia del método para aproximar los verdaderos valores de los datos faltantes o incongruentes. Para resumir los resultados, se presentan más adelante los valores producidos por el método computacional, que incorpora tanto la edición de datos incongruentes, como la imputación de los datos faltantes. El experimento que se diseñó, pretende reproducir los casos que podrían presentarse con mayor frecuencia en la práctica, los cuales se describen en el Cuadro 5.4. En ese cuadro se adoptó la notación siguiente, O = dato observado, - = dato con valor 25% menor que el observado y + = dato con valor 25% mayor que el observado.

Cuadro 5.4 Diseño experimental para las simulaciones

Corrida experimental	Valores de las variables			Corrida experimental	Valores de las variables			Corrida experimental	Valores de las variables		
	NC	PP	PC		NC	PP	PC		NC	PP	PC
1	O	O	O	11	O	ND	-	21	-	-	-
2	ND	O	O	12	O	ND	+	22	-	-	ND
3	-	O	O	13	O	ND	ND	23	-	ND	-
4	+	O	O	14	ND	-	-	24	-	ND	O
5	O	ND	O	15	ND	-	O	25	-	ND	ND
6	O	-	O	16	ND	+	O	26	+	+	+
7	O	+	O	17	ND	+	+	27	+	+	ND
8	O	O	ND	18	ND	ND	-	28	+	ND	+
9	O	O	-	19	ND	ND	O	29	+	ND	O
10	O	O	+	20	ND	ND	+	30	+	ND	ND

Aun sin realizar un análisis estadístico formal de los resultados del experimento, se aprecia con bastante claridad que los valores estimados por el procedimiento, ya sea para reemplazar datos incongruentes o para predecir datos faltantes, son razonables en tanto que la discrepancia de los pronósticos restringidos para valores faltantes, respecto

a los verdaderos valores de las variables, son cifras que por lo común fluctúan alrededor de $\pm 25\%$, que son las variaciones introducidas a propósito en los datos. De hecho, la discrepancia máxima es de -33.5% en PC, de 29.2% en PP, y de -28% en NC. Por otro lado, respecto a los valores del estadístico de compatibilidad, en principio se nota que los datos observados no son incompatibles con los pronósticos irrestrictos y que, al introducir los cambios indicados por el diseño muestral, en la mayor parte de los casos se produce incompatibilidad (con significancia menor o igual que 1%), principalmente cuando se incrementa o reduce el valor de NC, o cuando se incrementa el valor de PP.

Cuadro 5.5 Resultados de las simulaciones para el rastro X8, especie de ganado bovino
(Enero de 2004) Pronósticos irrestrictos: NC = 3350, PP = 1169862, PC = 680300

Restricciones			Est. de comp. (sig.)	Pronósticos restringidos			Observado - estimado en %		
NC	PP	PC		NC	PP	PC	NC	PP	PC
3401	1355800	744202	5.65 (0.13)	3401	1355800	744202	0.0	0.0	0.0
ND	1355800	744202	5.64 (0.13)	3390	1355800	744202	0.3	0.0	0.0
2551	1355800	744202	36.23 (0.00)	2551	1355800	744202	25.0	0.0	0.0
4251	1355800	744202	15.08 (0.00)	4251	1355800	744202	-25.0	0.0	0.0
3401	ND	744202	137.10 (0.00)	3401	1250250	744202	0.0	7.8	0.0
3401	1016850	744202	2.16 (0.34)	3401	1016850	578105	0.0	25.0	22.3
3401	1694750	744202	13.64 (0.00)	3401	1694750	945289	0.0	-25.0	-27.0
3401	1355800	ND	1.33 (0.51)	3401	1355800	784680	0.0	0.0	-5.4
3401	1355800	558152	1.33 (0.51)	3401	1355800	784680	0.0	0.0	-5.4
3401	1355800	930253	1.33 (0.51)	3401	1355800	784680	0.0	0.0	-5.4
3401	ND	558152	3.05 (0.22)	3401	997612	558152	0.0	26.4	25.0
3401	ND	930253	10.64 (0.00)	3401	1643031	930253	0.0	-21.2	-25.0
3401	ND	ND	0.04 (0.85)	3401	1182873	688499	0.0	12.8	7.5
ND	1016850	558152	14.73 (0.00)	2542	1016850	558152	25.3	25.0	25.0
ND	1016850	744202	12.88 (0.00)	2542	1016850	578080	25.3	25.0	22.3
ND	1694750	744202	11.91 (0.00)	4237	1694750	993150	-24.6	-25.0	-33.5
ND	1694750	930253	18.75 (0.00)	4237	1694750	930253	-24.6	-25.0	-25.0
ND	ND	558152	10.29 (0.01)	2611	976471	558152	23.2	28.0	25.0
ND	ND	744202	0.48 (0.79)	3482	1272588	744202	-2.4	6.1	0.0
ND	ND	930253	12.07 (0.00)	4352	1562918	930253	-28.0	-15.3	-25.0
2551	1016850	558152	14.43 (0.00)	2551	1016850	558152	25.0	25.0	25.0
2551	1016850	ND	12.54 (0.00)	2551	1016850	578356	25.0	25.0	22.3
2551	ND	558152	12.33 (0.00)	2551	978013	558152	25.0	27.9	25.0
2551	ND	744202	34.24 (0.00)	2551	1346147	744202	25.0	0.7	0.0
2551	ND	ND	12.30 (0.00)	2551	959650	547568	25.0	29.2	26.4
4251	1694750	930253	19.00 (0.00)	4251	1694750	930253	-25.0	-25.0	-25.0
4251	1694750	ND	12.11 (0.00)	4251	1694750	993774	-25.0	-25.0	-33.5
4251	ND	930253	1.42 (0.01)	4251	1565422	930253	-25.0	-15.5	-25.0
4251	ND	744202	10.04 (0.01)	4251	1290538	744202	-25.0	4.8	0.0
4251	ND	ND	9.40 (0.00)	4251	1395458	821452	-25.0	-2.9	-10.4

Cabe hacer notar que los pronósticos restringidos de las corridas experimentales 8, 9 y 10 son idénticos, tanto en el Cuadro 5.5 (NC = 3401, PP = 1355800 y PC = 784680) como en el Cuadro 5.6 (NC = 25, PP = 10000 y PC = 5190). La razón de esto es que en la corrida 8 se supone que el dato de PC es faltante y se estima con el procedimiento de imputación como es de esperarse, mientras que en las corridas 9 y 10, los valores que toma PC no satisfacen la restricción $K1 \cdot K3 \leq PC/NC \leq K2 \cdot K4$ y el proceso de edición inicial de los datos conduce a desechar los valores de PC como restricciones, de manera que sus datos e imputan como si fueran faltantes.

Cuadro 5.6 Resultados de las simulaciones para el rastro X2, especie de ganado bovino
(Enero de 2004) Pronósticos irrestrictos: NC = 7, PP = 2153, PP = 1234

Restricciones			Est. de comp. (sig.)	Pronósticos restringidos			Observado - estimado en %		
NC	PP	PC		NC	PP	PC	NC	PP	PC
25	10000	4500	6.94 (0.03)	25	10000	5190	0.0	0.0	-15.3
ND	10000	4500	6.94 (0.03)	25	10000	5190	0.0	0.0	-15.3
19	10000	4500	7.62 (0.02)	19	10000	5137	24.0	0.0	-14.2
31	10000	4500	8.30 (0.02)	31	10000	5446	-24.0	0.0	-21.0
25	ND	4500	5.97 (0.05)	25	7595	4500	0.0	24.1	0.0
25	7500	4500	5.81 (0.05)	25	7500	4157	0.0	25.0	7.6
25	12500	4500	8.91 (0.01)	25	12500	6296	0.0	-25.0	-39.9
25	10000	ND	6.94 (0.03)	25	10000	5190	0.0	0.0	-15.3
25	10000	3375	6.94 (0.03)	25	10000	5190	0.0	0.0	-15.3
25	10000	5625	6.94 (0.03)	25	10000	5190	0.0	0.0	-15.3
25	ND	3375	5.66 (0.06)	25	6413	3375	0.0	35.9	25.0
25	ND	5625	7.04 (0.03)	25	9385	5625	0.0	6.2	-25.0
25	ND	ND	5.51 (0.02)	25	6615	3615	0.0	33.9	19.7
ND	7500	3375	4.39 (0.11)	19	7500	3925	24.0	25.0	12.8
ND	7500	4500	4.39 (0.11)	19	7500	3925	24.0	25.0	12.8
ND	12500	4500	9.36 (0.01)	31	12500	6446	-24.0	-25.0	-43.2
ND	12500	5625	9.36 (0.01)	31	12500	6446	-24.0	-25.0	-43.2
ND	ND	3375	2.98 (0.23)	16	5659	3375	36.0	43.4	25.0
ND	ND	4500	5.13 (0.08)	21	7501	4500	16.0	25.0	0.0
ND	ND	5625	7.24 (0.03)	26	9334	5625	-4.0	6.7	-25.0
19	7500	3375	4.44 (0.11)	19	7500	3938	24.0	25.0	12.5
19	7500	ND	4.44 (0.11)	19	7500	3938	24.0	25.0	12.5
19	ND	3375	3.60 (0.17)	19	5738	3375	24.0	42.6	25.0
19	ND	4500	4.93 (0.09)	19	7616	4500	24.0	23.8	0.0
19	ND	ND	3.32 (0.07)	19	5110	2826	24.0	48.9	37.2
31	12500	5625	9.31 (0.01)	31	12500	6434	-24.0	-25.0	-43.0
31	12500	ND	9.31 (0.01)	31	12500	6434	-24.0	-25.0	-43.0
31	ND	5625	8.24 (0.02)	31	9446	5625	-24.0	5.5	-25.0
31	ND	4500	7.63 (0.02)	31	8208	4500	-24.0	17.9	0.0
31	ND	ND	7.62 (0.01)	31	8118	4396	-24.0	18.8	2.3

En el Cuadro 5.6 se observa que las mayores discrepancias porcentuales son: de 36% en NC, de 48.9% en PP, y de -43.2% en PC. Adicionalmente, los valores del estadístico de compatibilidad muestran que los datos observados son incompatibles con los pronósticos irrestrictos al nivel de significancia del 3%, lo cual se debe sin duda a

que los valores observados de las variables ($NC = 25$, $PP = 10000$ y $PC = 5190$) son de magnitudes casi 4 veces las de los valores esperados ($NC = 7$, $PP = 2153$ y $PC = 1234$). Al introducir los aumentos y reducciones porcentuales indicados por el diseño muestral, se mantiene la incompatibilidad al nivel de significancia menor o igual que 5%, en la mayor parte de los casos. Sin embargo, la hipótesis de compatibilidad tiende a no rechazarse (con significancia menor o igual que 6%), cuando los valores de NC y PP se desconocen o disminuyen, y PC no aumenta su valor.

6. CONSIDERACIONES FINALES

A manera de conclusión del presente estudio, a continuación se presentan algunos comentarios y recomendaciones que pretenden servir de guía para una correcta aplicación de la metodología propuesta. Debe subrayarse que el procedimiento consiste de dos fases, una primera que es de edición de los datos observados para las tres variables y que se realiza con el fin de que las nuevas observaciones puedan considerarse como datos válidos al momento de volver a estimar el modelo VAR correspondiente. Ya con el modelo reestimado, incluyendo los nuevos datos editados, se pueden obtener los valores imputados para los datos faltantes.

La fase de edición tiene en cuenta de manera fundamental la existencia de constantes que restringen de forma estructural el recorrido factible de las variables. El cumplimiento de las restricciones se vuelve forzoso a través de la aplicación de transformaciones a las variables. Sólo en el caso de la variable PC se permite que no se cumplan estrictamente las restricciones en los datos imputados; sin embargo, al ingresar esos datos al proceso de reestimación del modelo, la edición inicial los fuerza a cumplir con la restricción estructural. La idea de permitir que los datos imputados de PC no satisfagan exactamente la restricción estructural surgió debido a que no hay razón estricta para considerar a las constantes empleadas en las restricciones estructurales como rígidas, sino como sugerencias de límites de tolerancia. En la aplicación aquí reportada se pudo hacer uso de tales límites porque existían las constantes de manera explícita, pero seguramente ese no será el caso en otras aplicaciones del método.

Por su lado, la imputación de datos propuesta es básicamente resultado del empleo del método de pronósticos restringidos basado en un modelo VAR y ésta es la aportación más importante que hace este trabajo a la literatura sobre el tema. Debido a que dicho modelo estadístico se estima con los datos disponibles, es conveniente volver

a estimar el modelo cada vez que se tienen nuevos datos disponibles, lo cual ocurre en la ESGRM cada mes. Asimismo, es de esperar que conforme más datos se vayan acumulando, el modelo requiera ser revisado, particularmente en su estructura de retrasos, y esto se sugiere que se lleve a cabo anualmente.

Debe tenerse en mente que es posible que al acumular mayor cantidad de datos, los que sean más añejos proporcionen poca información acerca del futuro cercano de la serie. Por tal motivo, se considera razonable mantener una ventana fija del número de años, al hacer la revisión anual, para evitar trabajar con series demasiado largas, las cuales de otra manera podrían incluir efectos asociados con cambios estructurales ocurridos al inicio del periodo muestral. En México es comúnmente aceptado que existe un ciclo sexenal en la economía, por lo cual la recomendación para el sistema de imputación de la ESGRM es que la ventana sea de longitud seis años. Una ventaja adicional de utilizar un modelo estadístico de manera explícita, en lugar de un procedimiento basado en algún algoritmo para imputar datos, es que se puede medir la incertidumbre asociada con las imputaciones y se puede realizar inferencia, como la relacionada con el estadístico de compatibilidad. Esto ocurre a nivel de rastro, por especie y por variable en consideración. No obstante, es complicado calcular la incertidumbre al agregar los resultados para todos los rastros, pero sería deseable contar con esa medida, para darla a conocer a los usuarios “expertos” que pudieran sacar provecho de ella al realizar análisis estadístico de la ESGRM. Por ahora, esto se considera como una parte de la investigación futura por realizar sobre el tema.

Agradecimientos

Víctor M. Guerrero desea agradecer el apoyo brindado por las autoridades de la Dirección General de Contabilidad Nacional y Estadísticas Económicas del INEGI, para la realización de este trabajo. Asimismo desea hacer patente su agradecimiento a la Asociación Mexicana de Cultura, A. C., por el apoyo brindado a través de la Cátedra de Análisis de Series de Tiempo y Pronósticos en Econometría del ITAM, para desarrollar labores de investigación.

APÉNDICE. DEMOSTRACIÓN DE LAS EXPRESIONES (3.12) A (3.14).

Las ecuaciones (3.9) y (3.11) pueden escribirse de manera conjunta como

$$\begin{pmatrix} E(\nabla \mathbf{Z}_{N+1} | \mathbf{Z}) \\ \mathbf{Y} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_k \\ \mathbf{C} \end{pmatrix} \nabla \mathbf{Z}_{N+1} + \begin{pmatrix} -\mathbf{a}_{N+1} \\ \mathbf{u} \end{pmatrix}$$

con

$$E\begin{pmatrix} -\mathbf{a}_{N+1} \\ \mathbf{u} \end{pmatrix} | \mathbf{Z} = \mathbf{0}_{k+M} \text{ y } \text{Var}\begin{pmatrix} -\mathbf{a}_{N+1} \\ \mathbf{u} \end{pmatrix} | \mathbf{Z} = \begin{pmatrix} \Sigma_a & 0 \\ 0 & \Sigma_u \end{pmatrix}.$$

Por lo tanto, el método de Mínimos Cuadrados Generalizados produce el predictor lineal con ECM mínimo, que está dado por

$$\begin{aligned} \hat{\mathbf{Z}}_{N+1} &= [(\mathbf{I}_k \ C') \begin{pmatrix} \Sigma_a^{-1} & 0 \\ 0 & \Sigma_u^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{I}_k \\ \mathbf{C} \end{pmatrix}]^{-1} (\mathbf{I}_k \ C') \begin{pmatrix} \Sigma_a^{-1} & 0 \\ 0 & \Sigma_u^{-1} \end{pmatrix} \begin{pmatrix} E(\mathbf{Z}_{N+1} | \mathbf{Z}) \\ \mathbf{Y} \end{pmatrix} \\ &= [(\Sigma_a^{-1} \ C' \Sigma_u^{-1}) \begin{pmatrix} \mathbf{I}_k \\ \mathbf{C} \end{pmatrix}]^{-1} (\Sigma_a^{-1} \ C' \Sigma_u^{-1}) \begin{pmatrix} E(\mathbf{Z}_{N+1} | \mathbf{Z}) \\ \mathbf{Y} \end{pmatrix} \\ &= (\Sigma_a^{-1} + C' \Sigma_u^{-1} C)^{-1} [\Sigma_a^{-1} E(\mathbf{Z}_{N+1} | \mathbf{Z}) + C' \Sigma_u^{-1} \mathbf{Y}] \end{aligned}$$

con

$$\text{ECM}(\hat{\mathbf{Z}}_{N+1}) = (\Sigma_a^{-1} + C' \Sigma_u^{-1} C)^{-1}.$$

Ahora, el Lema de la Matriz Inversa establece que, si E y G son matrices no-singulares,

$$(E + FGH)^{-1} = E^{-1} - E^{-1}F(G^{-1} + HE^{-1}F)^{-1}HE^{-1}.$$

Este resultado permite deducir que

$$\begin{aligned} (\Sigma_a^{-1} + C' \Sigma_u^{-1} C)^{-1} &= \Sigma_a - \Sigma_a C' (\Sigma_u + C \Sigma_a C')^{-1} C \Sigma_a \\ &= (\mathbf{I}_k - \mathbf{A}_u C) \Sigma_a \end{aligned}$$

y

$$\begin{aligned} (\Sigma_a^{-1} + C' \Sigma_u^{-1} C)^{-1} C' \Sigma_u^{-1} &= [\Sigma_a^{-1} + (C' \Sigma_u^{-1} C) \Sigma_a \Sigma_a^{-1}]^{-1} C' \Sigma_u^{-1} \\ &= \{\Sigma_a - \Sigma_a (C' \Sigma_u^{-1} C) [\Sigma_a^{-1} + \Sigma_a^{-1} \Sigma_a (C' \Sigma_u^{-1} C)]^{-1} \Sigma_a^{-1} \Sigma_a\} C' \Sigma_u^{-1} \\ &= [\Sigma_a - \Sigma_a C' \Sigma_u^{-1} C (\Sigma_a^{-1} + C' \Sigma_u^{-1} C)^{-1}] C' \Sigma_u^{-1} \\ &= \Sigma_a C' \Sigma_u^{-1} - \Sigma_a C' \Sigma_u^{-1} C (\Sigma_a^{-1} + C' \Sigma_u^{-1} C)^{-1} C' \Sigma_u^{-1} \\ &= \Sigma_a C' [\Sigma_u^{-1} - \Sigma_u^{-1} C (\Sigma_a^{-1} + C' \Sigma_u^{-1} C)^{-1} C' \Sigma_u^{-1}] \\ &= \Sigma_a C' (\Sigma_u + C \Sigma_a C')^{-1} \\ &= \mathbf{A}_u. \end{aligned}$$

Por lo tanto,

$$\begin{aligned} \hat{\mathbf{Z}}_{N+1} &= (\mathbf{I}_k - \mathbf{A}_u C) E(\mathbf{Z}_{N+1} | \mathbf{Z}) + \mathbf{A}_u \mathbf{Y} \\ &= E(\mathbf{Z}_{N+1} | \mathbf{Z}) + \mathbf{A}_u [\mathbf{Y} - C E(\mathbf{Z}_{N+1} | \mathbf{Z})] \end{aligned}$$

con

$$ECM(\hat{\mathbf{Z}}_{N+1}) = (\mathbf{I}_k - \mathbf{A}_u \mathbf{C}) \Sigma_a.$$

Referencias

- Caporello, G. y Maravall, A. (2002). "A tool for quality control of time series data. Program TERROR". Disponible en el sitio del Banco de España en la Web, <http://www.bde.es/servicio/software/econom.htm>
- Engle, R. F. y Granger, C. W. J. (1987). "Co-integration and error correction representation, estimation and testing". *Econometrica* **55**, 251 - 276.
- Ganaderos (2005). www.mexicoganadero.com/limousin
- Guerrero, V. M. (2003). *Análisis Estadístico de Series de Tiempo Económicas*, 2ª. edición. Thomson-Learning, México.
- Guerrero, V. M., Pena, B., Senra, E. y Alegría, A. (2005). "Restricted forecasting with a VEC model: Validating the feasibility of economic targets". Disponible en la página Web de Eva Senra <http://www2.uah.es/>.
- INEGI (2004a). *Síntesis metodológica de la Estadística de Sacrificio de Ganado en Rastros Municipales*. <http://www.inegi.gob.mx/>
- INEGI (2004b). *Publicación anual de la Estadística de Sacrificio de Ganado en Rastros Municipales por Entidad Federativa 1998 - 2003*. <http://www.inegi.gob.mx/>
- Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlín.
- Pankratz, A. (1989). "Time series forecasts and extra-model information". *Journal of Forecasting* **8**, 75 – 83.
- Pfefferman, D. y Nathan, G. (2002). "Imputation for Wave Nonresponse: Existing Methods and a Time Series Approach". En Groves, R. M., Dillman, D. A., Eltinge, J. L. y Little, R. J. A. (eds.) *Survey Nonresponse*. John Wiley and Sons, New York. 417 – 429.
- Ruíz F. A., Cavallotti V. B., Villegas G. A., Ramírez S. F. y Armendáriz M. J. (2001). "Diagnóstico del sector pecuario". En: *México Rural: Políticas para su reconstrucción*. Universidad Autónoma Chapingo, Chapingo, México.
- SAGARPA (2005). www.sagarpa.gob.mx/ganaderito
- Sims, C.A. (1980). "Macroeconomics and Reality". *Econometrica* **48**, 1 - 48.
- Veterinaria (2005). www.veterin.unam.mx