

Aplicaciones en Economía y Ciencias Sociales con Stata

Alfonso Mendoza Velázquez, editor de la colección
Centro de Investigación e Inteligencia Económica CIIE-UPAEP



Una Publicación de Stata Press
StataCorp LP
College Station, Texas



Derechos de autor © 2013 StataCorp LP
Todos los derechos reservados. Primera Edición 2013

Publicado por Stata Press, 4905 Lakeway Drive, College Station, Texas 77845

Composición tipográfica en L^AT_EX 2_ε

Impreso en los Estados Unidos de América

10 9 8 7 6 5 4 3 2 1

ISBN-10: 1-59718-134-X

ISBN-13: 978-1-59718-134-1

Número de Control de la Librería del Congreso: 2013933553

Está prohibida la reproducción parcial o total de este libro. No está permitido que el material contenido en este libro sea guardado en un sistema de recuperación o sea transcrito por ningún medio—electrónico, mecánico, fotocopiado, grabado, o cualquier otro medio—sin la previa autorización por escrito de StataCorp LP.

Stata, **stata**, Stata Press, Mata, **mata**, and NetCourse son marcas registradas de StataCorp LP.

Stata and Stata Press son marcas registradas en la Organización Mundial de la Propiedad Intelectual de las Naciones Unidas.

L^AT_EX 2_ε es una marca registrada de la Sociedad Americana de Matemática.

Índice general

Índice de cuadros	XV
Índice de gráficas	XVII
Prefacio	XIX
Agradecimientos	XXIII
Notación y otras convenciones	XXV
I. Pobreza, desigualdad y valoración contingente	1
1. iop - Estimar desigualdad de oportunidades cuando el indicador es binario	3
1.1. Introducción	3
1.2. Metodología	4
1.2.1. Descomposición en fuentes de desigualdad	5
1.2.2. Descomposición al estilo Oaxaca-Blinder	5
1.3. Datos	6
1.4. El comando iop	7
1.4.1. Sintaxis de iop	7
1.4.2. Ejemplos	8
1.5. Resultados	9
1.5.1. Descomposición en fuentes de desigualdad	10
1.5.2. Descomposición al estilo Oaxaca-Blinder	10
1.6. Conclusiones	12
1.7. Ejercicios	12
1.8. Apéndice. Estimaciones probit	13

2.	Un modelo de valla doble para datos de conteo y su aplicación en el estudio de la fecundidad en México	15
2.1.	Introducción	15
2.2.	Datos y definición de las variables	16
2.3.	Aspectos econométricos	22
2.3.1.	Un modelo de valla doble	22
2.3.2.	Heterogeneidad no observada	26
2.3.3.	Relación con la literatura	28
2.4.	Sintaxis de los comandos empleados	29
2.5.	Resultados	31
2.5.1.	Modelo de valla simple	31
2.5.2.	Resultados de modelos de valla doble	35
	Ventajas del modelo de valla doble	39
	Efecto de las variables explicativas	43
2.6.	Conclusiones	45
2.7.	Ejercicios	46
3.	Inflación y pobreza en México (1993-2009)	47
3.1.	Introducción	47
3.2.	Revisión de la literatura	48
3.3.	Datos	49
3.3.1.	Medición de la pobreza	50
3.3.2.	Homologación de la ENEU y de la ENOE	52
3.3.3.	Análisis descriptivo	53
3.4.	Modelo econométrico	57
3.5.	Resultados	58
3.6.	Conclusiones	60
3.7.	Ejercicios	61
3.8.	Apéndice. Descripción de do-files	62
4.	Introducción a la valoración contingente utilizando Stata	75
4.1.	Introducción	75

<i>Índice general</i>	VII
4.2. Valoración contingente	76
4.2.1. Valoración contingente utilizando preguntas dicotómicas	77
4.2.2. Estimación econométrica del modelo dicotómico	79
4.2.3. Ejemplo del modelo dicotómico utilizando Stata	81
4.2.4. Valoración contingente utilizando preguntas dicotómicas con seguimiento	85
4.2.5. Modelo econométrico de datos por intervalos	86
4.2.6. Ejemplo del modelo dicotómico con seguimiento utilizando Stata	89
4.3. Ejercicios	92
II. Modelación macroeconómica	93
5. Choques transitorios y de largo plazo en el PIB mexicano: un modelo de vectores autorregresivos estructurales con Stata	95
5.1. Introducción	95
5.2. Vector autorregresivo estándar bivariado (VAR)	97
5.3. Identificación del modelo VAR: descomposición de Choleski	99
5.3.1. Funciones de impulso-respuesta	101
5.4. La descomposición de Blanchard y Quah	102
5.5. Aplicación en Stata	105
5.5.1. Estacionariedad de las series y pruebas de raíces unitarias	105
5.5.2. Vectores autorregresivos (VAR)	110
Funciones de impulso-respuesta	115
El comando IRF	117
El impacto de la crisis: los multiplicadores dinámicos	122
5.5.3. Vectores autorregresivos estructurales (VARs)	125
Restricciones de corto plazo: descomposición de Choleski	126
Método de Blanchard y Quah (identificación de largo plazo)	130
Descomposición de varianza	132
5.6. Conclusión	134
5.7. Ejercicios	135

6.	Prospectos para la economía: Una aplicación con modelos VAR cointegrados y proyecciones probabilísticas	137
6.1.	Introducción	137
6.2.	Economías y modelación	138
6.2.1.	Uruguay y Venezuela	138
6.2.2.	Modelación	144
6.3.	Modelo econométrico	146
	Identificación de las relaciones de largo plazo	147
	Estimación de los parámetros de corto plazo	148
	Proyecciones probabilísticas	148
	Enfoque paramétrico para las simulaciones	149
	Enfoque no paramétrico	149
6.4.	Estimaciones para Uruguay	150
6.4.1.	Selección del número de rezagos y tests de cointegración . .	150
6.4.2.	Modelo VAR cointegrado	152
6.4.3.	Proyecciones probabilísticas	154
	Simulaciones no paramétricas	155
	Escenario inercial	155
	Escenario de impacto moderado	157
6.5.	Estimaciones para Venezuela	159
	Grado de integración de las variables	159
6.5.1.	Selección del número de rezagos y tests de cointegración . .	160
6.5.2.	Modelo VAR cointegrado	162
6.5.3.	Proyecciones probabilísticas	164
	Escenario inercial	165
	Escenario de impacto extremo	167
6.6.	Conclusiones	169
6.7.	Ejercicios	170

III. Análisis electorales	173
7. Análisis de la elección presidencial de 2006 en México utilizando Stata	175
7.1. Introducción	175
7.2. Breve panorama de la elección de 2006	175
7.3. El problema analizado	177
7.3.1. Los datos	177
7.4. Análisis de las preferencias en la elección de 2006	177
7.4.1. La volatilidad de los electores	180
7.4.2. De la primera a la segunda ronda	181
7.4.3. De la segunda a la tercera ronda	183
7.5. Factores que motivaron el cambio de preferencias	185
7.5.1. Variables	186
Preferencias de la segunda ronda (variables dependientes)	186
Aspectos socioeconómicos, ideológicos y de entorno (variables independientes)	187
7.5.2. Estadística descriptiva	188
7.5.3. El modelo logit	193
7.5.4. Análisis inferencial de la elección en 2006 en México	194
Estimación de β con el modelo de probabilidad lineal	195
Estimación del vector δ con el modelo logit	197
Efectos marginales de las variables independientes para un modelo logit	199
Bondad de ajuste para un modelo logit	203
Matriz de clasificación	204
7.6. Conclusiones	205
7.7. Ejercicios	206
8. Modelos estadísticos para sistemas electorales multipartidistas en Stata	209
8.1. Introducción	209

8.2.	Estimación de modelos estadísticos con datos multipartidistas	211
8.2.1.	Mínimos cuadrados ordinarios	211
8.2.2.	Datos composicionales	214
8.2.3.	SURE	218
8.3.	Cantidades de interés	220
8.4.	Medidas de incertidumbre	225
8.4.1.	Simulación postestimación	226
8.4.2.	Medidas de incertidumbre asociadas a los sistemas electorales	228
8.5.	Conclusiones	233
9.	Métodos de imputación múltiple para predecir resultados electorales	235
9.1.	Introducción	235
9.2.	Marco teórico	236
9.2.1.	Teoría de la identificación partidaria	237
9.2.2.	Teoría del votante racional	238
9.2.3.	Teoría socio-estructural del voto	240
9.3.	Datos y metodología	241
9.3.1.	Datos	241
9.3.2.	VARIABLES	242
9.3.3.	Métodos muestrales de estimación	247
9.3.4.	Modelos de imputación	248
9.3.5.	Evaluación	250
9.4.	Resultados	251
9.4.1.	Capacidad predictiva de la imputación	252
9.4.2.	Los determinantes del voto	256
9.5.	Conclusiones	259
9.6.	Ejercicios	260
9.7.	Apéndice A. Diagnósticos de la imputación	260
9.8.	Apéndice B. Imputación múltiple con Stata	263

IV. Anexos	275
A. Introducción al manejo de bases de datos con Stata	277
A.1. Introducción	277
A.1.1. Inicio en Stata	277
A.1.2. El lenguaje de Stata	278
A.1.3. Presentación de resultados	279
A.1.4. Mensajes de error	280
A.1.5. Ayuda sobre comandos y recursos de Internet	280
A.1.6. Estimaciones con calificadores o restricciones en la muestra	281
A.1.7. Rutinas en Stata	283
A.2. Lectura, importación y exportación de bases de datos	284
A.2.1. Bases de datos en Stata	284
A.2.2. Importación de bases de datos	284
A.2.3. Guardar una base de datos	286
A.2.4. Exportación de bases de datos	286
A.2.5. Revisión y edición de datos	286
A.3. Manipulación de variables	287
A.3.1. Escalas de variables	287
A.3.2. Información detallada de una variable	287
A.3.3. Cambiar formato de variables numéricas	289
A.3.4. Cambiar el formato de una variable de fecha	289
A.3.5. Eliminar y renombrar variables	290
A.3.6. Etiquetas	291
Etiquetas de bases de datos	291
Etiquetas de variables	291
Etiquetas de valores	291
A.3.7. Notas	293
A.3.8. Modificación y creación de variables	294
A.3.9. Acciones repetidas para manipular variables	295
A.3.10. Colapsar variables	295

A.3.11.	Generación de variables dicotómicas	296
A.3.12.	Datos faltantes y valores atípicos	298
A.3.13.	Trabajo con datos faltantes	298
A.3.14.	Revisar datos faltantes	299
A.3.15.	Revisar datos atípicos	302
A.4.	Tipos de bases de datos	303
A.4.1.	Utilización de series de tiempo	305
A.4.2.	Rezagos y diferencias	305
A.4.3.	Trabajo con datos panel	307
Cambio de forma de una base de datos panel	307	
A.4.4.	Combinar bases de datos	310
A.5.	Ejercicios	312
B.	Métodos básicos de inferencia estadística y análisis de regresión	315
B.1.	Prueba de hipótesis	316
B.1.1.	El estadístico de prueba <i>vs.</i> el valor crítico	317
Propiedades de los estimadores	317	
B.1.2.	Estadísticos de prueba y valores críticos	318
B.1.3.	Pruebas de hipótesis en Stata	324
Nivel de significación empírico (p-value)	325	
Valores críticos	327	
Pruebas de hipótesis con varianza desconocida	327	
Proporciones	329	
Diferencia de medias	330	
Diferencia de dos proporciones	332	
Hipótesis sobre varianzas	334	
Coeficiente de correlación	337	
B.2.	Intervalos de confianza	338
B.2.1.	Intervalos de confianza y prueba de hipótesis	340
B.2.2.	Tamaño de la muestra	341

B.3.	Análisis de regresión	342
B.3.1.	Supuestos del modelo de regresión lineal clásico	342
B.3.2.	Estimación del modelo de regresión lineal	343
B.3.3.	Ajuste del modelo	344
B.3.4.	Análisis de varianza	346
B.3.5.	Inferencia sobre los parámetros de regresión	347
B.3.6.	Intervalos de confianza	349
B.3.7.	Regresión simple usando Stata	350
	Análisis de varianza y ajuste del modelo	352
	Estimadores de mínimos cuadrados	352
	Significancia estadística de los estimadores MCO	353
	Intervalos de confianza	353
B.4.	Comentarios finales	354
B.5.	Preguntas y ejercicios	354
	Referencias	357
	Índice de autores	371
	Índice temático	377

(Pages omitted)

Prefacio

Este libro integra aportes académicos de investigadores experimentados de la economía y las ciencias sociales. Es el primer libro publicado por Stata Press con un enfoque iberoamericano, y permitirá a los lectores en español conocer un conjunto de técnicas y desarrollos novedosos en su propio idioma. La obra tiene dos objetivos: primero, mostrar resultados de investigación originales en las ciencias sociales, y segundo, que el lector comprenda y aplique las técnicas de Stata 12.0. Cada capítulo ilustra un problema económico o social particular y demuestra cómo analizarlo y resolverlo gradualmente empleando menús y comandos. Ésta es una obra que sirve de apoyo al lector interesado en conocer más sobre temas económicos y sociales, pero que también busca aplicar el análisis teórico-estadístico que aquí se propone a su propio contexto. El libro mantiene así un equilibrio entre la teoría, el modelaje técnico, la programación y el análisis de datos.

La obra reúne nueve capítulos de investigación originales que contribuyen a la literatura académica en tres áreas: 1) pobreza, desigualdad y valuación contingente; 2) modelaje macroeconómico, y 3) análisis electoral. En cada una de estas áreas se describen y emplean diversas funciones estadísticas, comandos, macros y rutinas disponibles en el propio programa, pero también se presentan códigos de usuario originales, creados especialmente por los autores para este libro. Cada autor desarrolla formalmente un tema concreto en el área de su especialidad y concluye con una serie de ejercicios con el fin de reforzar la comprensión teórica y práctica. La serie de contribuciones de investigación concluye con dos capítulos adicionales, en forma de anexos, que buscan familiarizar a los usuarios con las técnicas de Stata y son un repaso de conceptos estadísticos básicos. Por lo tanto, esta obra constituye una fuente de referencia muy útil para el estudio de temas teóricos concretos, pero también de diversos comandos y rutinas en el ambiente Stata 12.0.

El libro surge de la necesidad de brindar a investigadores y estudiantes de licenciatura y posgrado una referencia formal pero al mismo tiempo accesible. Está dirigido a investigadores y estudiantes de economía, ciencia política, relaciones internacionales, ciencias de la salud y, en general, estudiantes que requieran elaborar una tesis con un soporte cuantitativo y marco de referencia adecuados. Los capítulos se distinguen por ilustrar, no sólo el uso y aplicabilidad de los comandos de Stata 12.0, sino mostrar cómo se usan de manera sistemática y lógica para resolver un problema de investigación concreto. Debido a esto, la obra puede emplearse incluso como libro de texto en cursos avanzados de econometría aplicada y como referencia en cursos de metodología de la investigación, para ilustrar la aplicación a distintos temas de interés actual.

Suponemos que el estudiante interesado ha llevado cursos introductorios de métodos cuantitativos en las ciencias sociales (álgebra básica, estadística descriptiva e inferencia y, de manera deseable, cálculo diferencial y econometría), así como un manejo básico de programas de estadística. Sin embargo, con el fin de repasar algunos de estos requerimientos, además de los capítulos de investigación el libro hace una revisión amplia de las características del programa Stata 12.0 y también un repaso breve de algunos conceptos estadísticos fundamentales. Asimismo, en cada capítulo los autores parten de conceptos básicos hasta alcanzar niveles intermedios, incluso avanzados, y proveen referencias que permiten a los lectores interesados profundizar más en los temas.

El libro está dividido en tres apartados. El primero, denominado “Pobreza, desigualdad y valoración contingente”, es iniciado por la contribución de Isidro Soloaga y Florian Wendelspiess Chávez, quienes resaltan el uso del comando `iop` para la estimación de la desigualdad de oportunidades cuando el indicador es binario. Este comando, desarrollado por los autores, se utiliza para evaluar el programa gubernamental mexicano Oportunidades. Se examina el impacto del programa sobre la igualación de oportunidades, el rol de las circunstancias personales y el acceso a niveles de bienestar. En el segundo capítulo, Alfonso Miranda propone un innovador modelo de valla doble de conteo de Poisson, con una aplicación al estudio de los determinantes de la fecundidad en México. Utilizando la Encuesta Nacional de Dinámica Demográfica (Enadid) de 1997, el autor pone especial acento en el estudio del impacto de la religión y el grupo étnico sobre la probabilidad de transición de conteos bajos a conteos altos. Para llevar a cabo la estimación, desarrolla varias piezas de código en Stata y describe su implementación en este estudio. Los resultados indican que la educación y el catolicismo están asociados a la reducción de la probabilidad de transición de una familia con cuatro niños a conteos de orden superior. En contraste, hablar una lengua indígena aumenta la probabilidad de tener una familia numerosa. En el tercer capítulo, utilizando un modelo de efectos fijos, Carlo Alcaraz y Carlos Nakashima examinan el impacto de la inflación sobre la pobreza en México de 1993 a 2009 y presentan evidencia respecto a la asociación entre pobreza y empleo informal. También muestran didácticamente cómo obtener información detallada de ingresos laborales a partir de la Encuesta Nacional de Empleo Urbano (ENEU) y la Encuesta Nacional de Ocupación y Empleo (ENOE). En el cuarto capítulo, Alejandro López-Feldman muestra el uso de los comandos `singleb` y `doubleb` desarrollados por él mismo para la estimación de la disposición a pagar, particularmente cuando la variable dependiente es dicotómica; ilustra la aplicación de estos comandos tomando como ejemplo el Parque Natural Alentejo, de Portugal, e ilustra las alternativas de valuación tradicionales usando modelos Probit.

El segundo apartado trata problemas de modelación macroeconómica. En el quinto capítulo, Alfonso Mendoza y Peter N. Smith ilustran el uso de Stata 12.0 para el examen de la relación entre el crecimiento del PIB y la inflación en México. Estudian la respuesta del crecimiento del PIB y los precios a choques de oferta y demanda, así como el impacto de corto y de largo plazo de los choques durante la “Crisis del Tequila”, la crisis asiática y la crisis hipotecaria. Los comandos para el análisis de series de tiempo ayudan a determinar en este estudio las propiedades de estacionariedad, las funciones de impulso-respuesta, el análisis de varianza y, finalmente, la estimación de vectores au-

torregresivos estructurales. Los autores emplean como método de extracción de choques el enfoque de Blanchard y Quah (1989). En el sexto capítulo, combinando las herramientas de vectores autorregresivos cointegrados y proyecciones probabilísticas, Gustavo Sánchez y Harold Zarvace exploran la utilización de Stata para producir prospectos de desarrollo en las economías venezolana y uruguaya. La aplicación tiene relevancia para los sectores público y privado debido al interés general que siempre despierta la proyección a futuro del PIB y la inflación, y también como insumos para valorar los riesgos y las oportunidades en la toma de decisiones.

El análisis de fenómenos electorales se presenta en el tercer apartado de la obra. El séptimo capítulo examina la elección presidencial de 2006 en México, empleando herramientas de estadística descriptiva e inferencia para el análisis de acontecimientos políticos. Ignacio Ibarra emplea un modelo de probabilidad lineal y un modelo logit para investigar cómo cambiaron las preferencias de los electores a lo largo de esa campaña. El autor determina la importancia de los diversos factores que motivaron ese cambio en las preferencias, por ejemplo, los escándalos políticos. Entre las conclusiones destaca que la elección de 2006 en México no parece haber cambiado el *statu quo*, sino más bien fue una elección que permitió preservarlo. Enseguida, en el capítulo 8, Javier Márquez y Javier Aparicio proponen modelos estadísticos para el análisis de sistemas electorales multipartidistas, adecuados para regímenes políticos como los de América Latina, con características comunes, tales como la separación de poderes (Ejecutivo y Legislativo), la elección por representación proporcional y la existencia de más de dos partidos políticos relevantes (sistemas multipartidistas). Los autores resaltan la utilidad de los modelos estadísticos para explicar o predecir la conformación de la asamblea en sistemas multipartidistas; con este objetivo, desarrollan el módulo `camaradip` en Stata 12.0, el cual incorpora técnicas útiles para adecuar los modelos estadísticos convencionales al estudio de sistemas electorales multipartidistas. El capítulo muestra paso a paso los componentes del modelo estadístico, tomando como caso de estudio la elección de diputados federales de 2006 en México. En el último capítulo de esta obra, Modesto Escobar y Antonio M. Jaime continúan el análisis de los procesos electorales, pero ahora utilizando técnicas de imputación de datos, las cuales permiten extender el pronóstico electoral a individuos acerca de los cuales no se dispone de información completa. Se aprovecha la literatura sobre el tratamiento de datos incompletos con Stata para obtener predicciones de las decisiones electorales individuales, cuando la no respuesta sesga sistemáticamente los pronósticos. Por el lado teórico se retoman los micro-fundamentos de la decisión electoral para la obtención de pronósticos a nivel agregado. La capacidad predictiva de estas técnicas se evalúa con las elecciones generales de 2011 en España, utilizando las encuestas electorales del Centro de Investigaciones Sociológicas.

(Pages omitted)

- Si los elementos de \mathbf{y}_t son $I(1)$ y están cointegrados con $\text{rango}(\mathbf{\Pi}) = r$, entonces habrá r combinaciones lineales de \mathbf{y}_t , $\mathbf{ecm}_t = \boldsymbol{\beta}'\mathbf{y}_t$, las cuales serán $I(0)$, y representan las relaciones de cointegración del sistema. El número de relaciones de cointegración, r , será estrictamente menor que el número de variables incluidas en el modelo.

Identificación de las relaciones de largo plazo

Se requieren al menos r^2 restricciones para identificar los coeficientes contenidos en $\boldsymbol{\alpha}$ y $\boldsymbol{\beta}$. El método de Johansen (1988) establece restricciones “estadísticas”⁶ para estimar esos parámetros. Con base en esas restricciones, el estimador por máxima verosimilitud de Johansen para los vectores de cointegración en $\boldsymbol{\beta}$ corresponde a los primeros r autovectores de una matriz de correlación canónica. Sin embargo, el conjunto de restricciones estadísticas no necesariamente será consistente con la teoría económica, y no hay razones teóricas para suponer que los vectores de cointegración deban ser ortogonales como lo sugiere el conjunto de restricciones en el método de Johansen. Adicionalmente, cuando $r > 1$ las relaciones de cointegración basadas en el método de Johansen pueden carecer de interpretación económica.

El enfoque de Pesaran y Shin (1999) y de Pesaran, Shin y Smith (2000) plantea estimar las relaciones de cointegración basado en una estructura de largo plazo provista por teoría económica *a priori*. Garratt et al. (2006) muestra que se pueden expresar las restricciones sobre las relaciones de cointegración como:

$$\mathbf{Rvec}(\boldsymbol{\beta}^*) = \mathbf{B}$$

Donde:

- \mathbf{R} es una matriz $(k \times (m + 1)r)$ de rango fila completo.
- \mathbf{B} es un vector $(k \times 1)$ de constantes conocidas.
- $\mathbf{vec}(\boldsymbol{\beta}^*)$ es un vector $(m + 1)r$ que agrupa las columnas de $\boldsymbol{\beta}$ (una debajo de la otra) en una sola columna.

La estimación, entonces, se realiza a través de una generalización del método de máxima verosimilitud de Johansen, incorporando las restricciones teóricas para la identificación de los parámetros del modelo. Este proceso se inicia con el estimador exactamente identificado obtenido a través del método de Johansen, y luego se maximiza una función de Lagrange para tomar en cuenta las $k - r^2$ restricciones adicionales. A diferencia del método de Johansen, las restricciones iniciales para la identificación exacta están basadas en la teoría económica.

⁶Esas restricciones son referidas como “estadísticas” por Garratt et al. (2006) porque no están basadas en la teoría económica.

Estimación de los parámetros de corto plazo

Las estimaciones máximo verosímiles de los parámetros de corto plazo del modelo podrían obtenerse usando mínimos cuadrados ordinarios para la regresión de $\Delta \mathbf{y}_t$ sobre **ecm**, $\Delta \mathbf{y}_{t-1}, \dots, \Delta \mathbf{y}_{t-p+1}$, donde $\mathbf{ecm} = \hat{\boldsymbol{\beta}}^{*'} \mathbf{y}_{t-1}^*$. Realizando la estimación de esta manera, $\hat{\boldsymbol{\beta}}$ es superconsistente y los estimadores de los parámetros de corto plazo son consistentes. Además, si realizamos el ajuste por mínimos cuadrados ordinarios podrían aplicarse las pruebas de diagnóstico estándar sobre los supuestos del modelo lineal. No obstante, el estimador por máxima verosimilitud de Johansen utilizado en este estudio permite hacer la estimación conjunta de los parámetros de todas las ecuaciones del modelo, lo cual toma en cuenta la correlación contemporánea entre las innovaciones de las ecuaciones del sistema.

El comando de Stata `vec` implementa la estimación de Johansen, y además de permitirnos obtener predicciones dinámicas con el comando de postestimación `fcast compute`, también nos da la posibilidad de realizar análisis a través de las funciones de impulso-respuesta y la descomposición de la varianza de la predicción. En los ejercicios desarrollados en este capítulo utilizamos las predicciones dinámicas para combinarlas con las proyecciones probabilísticas definidas a continuación.

Proyecciones probabilísticas

Esta herramienta permite la obtención de probabilidades de ocurrencia de eventos condicionados a la información disponible en el periodo de estimación. Estas probabilidades pueden ser estimadas sobre la base de un modelo macroeconómico, y el evento puede ser definido con respecto a los valores de una variable, o de un conjunto de variables, medidas en un periodo particular de tiempo o a lo largo de una secuencia de periodos en el futuro.

Los ejercicios empíricos de este capítulo utilizan los parámetros estimados de modelos VEC para Uruguay y Venezuela. Se plantean eventos individuales y combinados respecto a la inflación y el producto interno bruto, y se realizan simulaciones para estimar las proyecciones probabilísticas acerca de niveles de inflación y crecimiento dentro de un rango asociado a escenarios de interés para las autoridades económicas. Por ejemplo, se pueden definir eventos como:

- E1 Crecimiento del producto de entre tres y cuatro por ciento.
- E2 Inflación inferior a seis por ciento
- E4 Crecimiento del producto de entre tres y cuatro por ciento, conjuntamente con una inflación inferior a seis por ciento

Este enfoque es atractivo porque, en lugar de producir proyecciones puntuales que pueden estar complementadas con intervalos de confianza para caracterizar la incertidumbre, se muestran proporciones que reflejan el número de veces que ocurre el evento en las simulaciones de las proyecciones. De esta manera, las autoridades económicas pueden contar con potenciales probabilidades de ocurrencia de los eventos de interés que sean definidos en esta parte del análisis.

A fin de calcular las probabilidades se deben simular por muestreo los pronósticos del modelo especificado. Esta simulación se puede realizar mediante dos diferentes enfoques: paramétrico y no paramétrico.

Enfoque paramétrico para las simulaciones

En este enfoque los errores son extraídos aleatoriamente de una distribución multivariada. Por ejemplo, si suponemos que los errores siguen una distribución normal:

$$\zeta_{T+i} \sim N(0, I)$$

se extraen m residuos aleatorios de esa distribución, y se transforman con base en la descomposición de Choleski de la matriz de varianzas y covarianzas de los errores del modelo, a fin de mantener la estructura generada por el ajuste del modelo VEC. Se suman entonces esos residuos simulados a las proyecciones del modelo y se obtienen de esta manera las proyecciones probabilísticas que servirán de base para analizar diferentes eventos sobre las variables de interés.

Enfoque no paramétrico

Esta alternativa corresponde a la selección de h extracciones aleatorias con reemplazamiento de los vectores residuales. En este caso, los errores simulados tendrán la misma distribución de la muestra original. Con los residuos obtenidos se sigue una metodología similar a la del método paramétrico, y se suman estos residuos simulados a las proyecciones del modelo para obtener las proyecciones probabilísticas.

El procedimiento básicamente consiste en utilizar los residuos históricos obtenidos del ajuste del modelo y tomar muestras (con reemplazamiento) repetidas para el número de periodos que serán proyectados.

Ahora bien, las muestras de los residuos históricos no pueden ser utilizadas directamente porque estarían significativamente influenciadas por la estructura temporal de las series. El enfoque desarrollado por Pesaran y Shin sugiere la construcción de una nueva serie de residuos basada en las proyecciones del modelo y que refleje la estructura de la distribución conjunta de esos residuos, pero que reduzca la dependencia temporal de los mismos. Esta nueva serie se obtiene multiplicando los residuos por la inversa de la descomposición de Choleski asociada a la matriz de varianzas y covarianzas de los residuos del modelo VAR cointegrado.

Se procede entonces a tomar las muestras repetidas para los periodos de proyección y se vuelven a transformar los vectores de residuos aleatorios utilizando esta vez la descomposición de Choleski (no la inversa) asociada a la matriz de varianzas y covarianzas para mantener la estructura de los residuos originales.

6.4. Estimaciones para Uruguay

El modelo para el caso uruguayo se estimó basado en las observaciones para el periodo comprendido entre el primer trimestre de 1989 y el segundo trimestre de 2011. Los datos fueron obtenidos de la página web del Banco Central de Uruguay y de las estadísticas financieras internacionales del Fondo Monetario Internacional.⁷ Los datos utilizados están disponibles en un archivo de Stata que puede ser bajado de la página web que contiene los archivos para este libro. A continuación, definimos las variables usadas en las estimaciones para Uruguay:

<code>m1</code>	Liquidez monetaria real.
<code>pib</code>	Producto interno bruto real.
<code>tipp906bn</code>	Tasa de interés pasiva para depósitos a 90 días, promedio puntual para los seis principales bancos.
<code>tcpn</code>	Tipo de cambio puntual.
<code>ipcp97</code>	Índice de precios al consumidor puntual (1997 = 100).
<code>mt</code>	Importaciones totales reales.
<code>xt</code>	Exportaciones totales reales.
<code>ipex</code>	Índice de precios de las exportaciones.

Un paso previo a la estimación de las relaciones de largo plazo y del modelo VAR cointegrado correspondió a la determinación del grado de integración de cada una de las variables endógenas, así como de la variable que será tratada como débilmente exógena (`ipex`). Basado en los tests de Dickey-Fuller (1979) aumentado y Phillips-Perron (1988), no se pudo rechazar la hipótesis nula de que todas las variables son integradas de orden 1 para el periodo muestral. La evaluación de los tests de raíces unitarias para todas las variables se incluye como ejercicio al final de este capítulo.

6.4.1. Selección del número de rezagos y tests de cointegración

El siguiente paso para la estimación del VAR cointegrado corresponde a la determinación del número de rezagos que serán incluidos para los tests de cointegración y para el ajuste del modelo. El *output* del comando `varsoc` muestra que, con la excepción de los criterios de Hanna-Quinn y Schwarz, los resultados sugieren un número de rezagos de alrededor de cuatro para el VAR subyacente. Tomando en cuenta estos resultados y la frecuencia trimestral de las series utilizadas en este estudio, se seleccionaron cuatro rezagos para el resto de los tests y las estimaciones.

⁷Véanse <http://www.bcu.gub.uy> y <http://www.imf.org/external/data.htm> (último acceso 1 de febrero de 2013).

```
. varsoc lm1 lpib ltipp906bn lipcp97 ltcpn lmt lxt lipex, maxlag(4)
Selection-order criteria
Sample: 1990q1 - 2011q2                Number of obs   =      86
```

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	267.063				3.3e-13	-6.02473	-5.93284	-5.79642
1	1126.87	1719.6	64	0.000	3.1e-21	-24.5319	-23.705	-22.4771
2	1281.89	310.04	64	0.000	3.8e-22	-26.6487	-25.0867	-22.7674*
3	1394.14	224.48	64	0.000	1.4e-22	-27.7706	-25.4735*	-22.0628
4	1481.91	175.55*	64	0.000	9.6e-23*	-28.3235*	-25.2913	-20.7892

```
Endogenous:  lm1 lpib ltipp906bn lipcp97 ltcpn lmt lxt lipex
Exogenous:   _cons
```

A continuación se requiere determinar el número de relaciones de cointegración existentes entre las variables del modelo. Usamos el comando `vecrank` para obtener los estadísticos de contraste para el test de Johansen correspondientes a la traza y al máximo autovalor. El estadístico de la traza (*trace statistic*) sugiere la posible presencia de hasta cinco o seis relaciones de cointegración, mientras que el estadístico correspondiente al máximo autovalor (*max statistic*) tiene un valor crítico muy cercano al 5% en el caso de la hipótesis nula para un número no mayor de tres relaciones de cointegración.

```
. vecrank lm1 lpib ltipp906bn lipcp97 ltcpn lmt lxt lipex, levela max lags(4)
Johansen tests for cointegration
Trend: constant                Number of obs =      86
Sample: 1990q1 - 2011q2                Lags =      4
```

maximum rank	parms	LL	eigenvalue	trace statistic	5% critical value	1% critical value
0	200	1342.3712		279.0761	156.00	168.36
1	215	1381.1684	0.59435	201.4816	124.24	133.57
2	228	1411.0026	0.50033	141.8133	94.15	103.18
3	239	1436.2588	0.44420	91.3008	68.52	76.07
4	248	1452.7831	0.31906	58.2522	47.21	54.46
5	255	1464.9991	0.24730	33.8202*1	29.68	35.65
6	260	1475.9395	0.22464	11.9395*5	15.41	20.04
7	263	1480.5188	0.10102	2.7809	3.76	6.65
8	264	1481.9092	0.03182			

maximum rank	parms	LL	eigenvalue	max statistic	5% critical value	1% critical value
0	200	1342.3712		77.5945	51.42	57.69
1	215	1381.1684	0.59435	59.6683	45.28	51.57
2	228	1411.0026	0.50033	50.5124	39.37	45.10
3	239	1436.2588	0.44420	33.0486	33.46	38.77
4	248	1452.7831	0.31906	24.4320	27.07	32.24
5	255	1464.9991	0.24730	21.8807	20.97	25.52
6	260	1475.9395	0.22464	9.1587	14.07	18.63
7	263	1480.5188	0.10102	2.7809	3.76	6.65
8	264	1481.9092	0.03182			

6.4.2. Modelo VAR cointegrado

De acuerdo con los resultados de los tests de cointegración, y con base en las relaciones presentadas en la sección 6.2.2, se ajustó un modelo VAR cointegrado con tres relaciones de largo plazo que fueron normalizadas en términos de ecuaciones para la demanda de dinero, para las importaciones y para las exportaciones.⁸ Usamos la siguiente especificación del comando `vec` para estimar los parámetros del modelo:

```

** Restricciones vec **
*** restricciones lm1 ***
  constraint 1 [_ce1]lm1=1
  constraint 2 [_ce1]lipcp97=0
  constraint 3 [_ce1]ltcpn=0
  constraint 4 [_ce1]lmt=0
  constraint 5 [_ce1]lxt=0
  constraint 6 [_ce1]lipex=0
*** restricciones lmt ***
  constraint 7 [_ce2]lm1=0
  constraint 8 [_ce2]ltipp906bn=0
  constraint 9 [_ce2]lipcp97=0
  constraint 10 [_ce2]lmt=1
  constraint 11 [_ce2]lxt=0
  constraint 12 [_ce2]lipex=0
*** restricciones lxt ***
  constraint 13 [_ce3]lm1=0
  constraint 14 [_ce3]lpib=0
  constraint 15 [_ce3]ltipp906bn=0
  constraint 16 [_ce3]lmt=0
  constraint 17 [_ce3]lxt=1
*** restricciones sobre alpha ***
  constraint 18 [D_lipex]1._ce1=0
  constraint 19 [D_lipex]1._ce2=0
  constraint 20 [D_lipex]1._ce3=0
  set more off

** Selected Equation **
  vec lm1 lpib ltipp906bn lipcp97          ///
      ltcpn lmt lxt lipex,                ///
      bconstraints(1/17)                   ///
      aconstraint(18/20)                   ///
      lags(4) rank(3) noetable             ///
      ltolerance(1e-7) tolerance(1e-4) noidtest

```

Las primeras seis restricciones corresponden a la relación de largo plazo para la demanda de dinero real; el coeficiente de `lm1` se restringe de manera que sea igual a uno para que la primera ecuación se pueda expresar en términos de `lm1` en función de las variables no restringidas `ltipp906bn` y `lpib`. Las siguientes seis restricciones corresponden a la ecuación de largo plazo de las importaciones, donde se expresa `lmt` en función del PIB y de la tasa de cambio nominal. El tercer grupo de restricciones determina la relación de largo plazo para las exportaciones, las cuales están afectadas por el índice de precios internos, la tasa de cambio nominal y el índice de precios de la canasta de exportaciones de Uruguay. Finalmente, se añaden tres restricciones que igualan a cero el efecto de cada una de las relaciones de cointegración sobre la

⁸La ecuación de paridad de intereses (6.7) no resultó significativa para ninguno de los dos países.

ecuación del VAR para el índice de precios de las exportaciones **ipex**. Estas últimas tres restricciones permiten tratar a la variable **ipex** como débilmente exógena.

El comando **vec** lista las variables contenidas en el modelo y luego especifica una serie de opciones, donde se incluyen las restricciones sobre los vectores de cointegración y sobre los parámetros asociados a la velocidad del ajuste hacia el equilibrio (o parámetros de impacto) en la ecuación para el **ipex**. También se indica el número de relaciones de cointegración con la opción **rank(3)** y el número de rezagos del modelo con la opción **lags(4)**.

Las otras dos opciones (**ltolerance** y **tolerance**) reducen el nivel de tolerancia para la optimización,⁹ y la opción **noidentest** indica que no se debe reportar el test de sobreidentificación.

Se presentan a continuación las estimaciones obtenidas con el comando **vec** especificado arriba:

```

Vector error-correction model
Sample: 1990q1 - 2011q2
Log likelihood = 1414.246
Det(Sigma_ml) = 7.19e-25
No. of obs = 86
AIC = -27.58711
HQIC = -24.96839
SBIC = -21.08023

Cointegrating equations
Equation      Parns    chi2      P>chi2
-----
_ce1           2      170.89    0.0000
_ce2           2     115.2276  0.0000
_ce3           3     111.1596  0.0000

Identification:  beta is underidentified
( 1)  [_ce1]lm1 = 1
( 2)  [_ce1]lipcp97 = 0
( 3)  [_ce1]ltcpn = 0
( 4)  [_ce1]lmt = 0
( 5)  [_ce1]lxt = 0
( 6)  [_ce1]lipex = 0
( 7)  [_ce2]lm1 = 0
( 8)  [_ce2]ltipp906bn = 0
( 9)  [_ce2]lipcp97 = 0
(10)  [_ce2]lmt = 1
(11)  [_ce2]lxt = 0
(12)  [_ce2]lipex = 0
(13)  [_ce3]lm1 = 0
(14)  [_ce3]lpib = 0
(15)  [_ce3]ltipp906bn = 0
(16)  [_ce3]lmt = 0
(17)  [_ce3]lxt = 1

```

⁹En ejercicios empíricos puede ser necesaria la reducción de los niveles de tolerancia a fin de alcanzar la convergencia. Sin embargo, estas modificaciones deben ser realizadas con cautela, evitando reducir de manera significativa la tolerancia.

beta	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_ce1						
lm1	1
lpib	-2.388095	.2667403	-8.95	0.000	-2.910897	-1.865294
ltipp906bn	.2394962	.0401545	5.96	0.000	.1607948	.3181976
lipcp97	0	(omitted)				
ltcpn	0	(omitted)				
lmt	0	(omitted)				
lxt	0	(omitted)				
lipex	0	(omitted)				
_cons	34.28762
_ce2						
lm1	0	(omitted)				
lpib	-1.190284	.1176426	-10.12	0.000	-1.420859	-.9597084
ltipp906bn	0	(omitted)				
lipcp97	0	(omitted)				
ltcpn	.1906142	.0365307	5.22	0.000	.1190154	.262213
lmt	1
lxt	0	(omitted)				
lipex	0	(omitted)				
_cons	4.087814
_ce3						
lm1	0	(omitted)				
lpib	0	(omitted)				
ltipp906bn	0	(omitted)				
lipcp97	268.7157	28.66662	9.37	0.000	212.5301	324.9012
ltcpn	-130.17	27.2982	-4.77	0.000	-183.6735	-76.66648
lmt	0	(omitted)				
lxt	1
lipex	-337.4693	45.183	-7.47	0.000	-426.0264	-248.9123
_cons	470.4702

La salida de Stata muestra los vectores de cointegración para cada una de las variables normalizadas. Estas relaciones de largo plazo se pueden expresar de la siguiente manera:

$$lm1 = 2.39 \times lpib - 0.24 \times ltipp906bn - 34.29$$

$$lmt = 1.19 \times lpib - 0.19 \times ltcpn - 4.09$$

$$lxt = -268.72 \times lipcp97 + 130.17 \times ltcpn + 337.47 \times lipex - 470.47$$

6.4.3. Proyecciones probabilísticas

En esta parte del análisis presentamos los resultados de algunos ejercicios de simulación que permiten producir probabilidades, en términos de distribuciones no paramétricas, para eventos sobre algunas de las variables del modelo. Estos ejercicios son de particular interés para las autoridades económicas debido a que no solamente se limitan a la presentación de proyecciones puntuales, sino que también pueden contestar preguntas sobre la posibilidad de que los niveles o las variaciones de una o más variables estén

por encima o por debajo de un determinado valor, e incluso la posibilidad de que esos niveles o variaciones de la variable estén dentro de un rango definido por el investigador.

Un aspecto adicional de los ejercicios de proyección que se presentan en esta sección es que se trabaja con dos tipos de escenarios. Un primer escenario “inercial”, donde se realizan predicciones dinámicas para cada una de las variables del modelo, y un segundo escenario de “impacto moderado”, en el cual se fijan niveles predeterminados para la variable débilmente exógena y el resto de las variables son proyectados de manera dinámica a partir de los parámetros del modelo.

Simulaciones no paramétricas

Luego de ajustar el modelo debemos preparar las proyecciones puntuales y los residuos para realizar las simulaciones. Los ejercicios que se presentan en esta sección están basados en la estimación para el periodo 1991Q1-2011Q2, y las proyecciones se realizan para el resto de 2011 y el año 2012. Por lo tanto, se toman muestras de seis observaciones de los residuos históricos para las simulaciones. Los residuos simulados son sumados a las proyecciones puntuales del modelo y de esta manera se obtienen las predicciones simuladas, que son entonces utilizadas para el análisis de los eventos asociados a las proyecciones.

Los comandos de Stata para implementar el procedimiento descrito en la sección 3, y en general para todos los ejercicios presentados en este capítulo, están contenidos en los “do-files” que pueden ser descargados del sitio web de Stata para este libro.

Escenario inercial

En este caso solamente se utilizan los resultados del modelo VAR sin tomar en cuenta posibles niveles prefijados para la variable exógena. El cuadro 6.5 muestra los resultados de las simulaciones para algunos eventos de interés en el caso del producto interno bruto de Uruguay. De acuerdo con las simulaciones del modelo es altamente probable que la variación del PIB de 2011 sea inferior a 6%, y 33% de las proyecciones simuladas corresponden a una variación del PIB que se ubicaría entre 5.5% y 7.5%. Para el año 2012 habría una expectativa de alrededor de 52% de que el PIB crezca en más de 5.9%, mientras que 30% sería la probabilidad de que el crecimiento esté entre 4.2 y 5.9 por ciento.

La segunda parte del cuadro 6.5 muestra un sumario de estadísticas descriptivas de las proyecciones simuladas, donde se observa que el promedio de las proyecciones simuladas para el crecimiento del PIB se ubica en 5.34 para el año 2011 y en 6.12 para el año 2012.