Datos de Panel

Guía para el uso de Eviews

Alfredo Baronio – Ana Vianco

Departamento de Matemática y Estadística
Facultad de Ciencias Económicas
Universidad Nacional de Río Cuarto
Noviembre de 2014

Contenido

INTRODUCCIÓN **3**

LA TÉCNICA DE DATOS DE PANEL **5**

ESPECIFICACIÓN GENERAL DE UN MODELO DE DATOS DE PANEL. 6

A) COMPONENTES DEL ERROR 7

B) VENTAJAS Y DESVENTAJAS DE LA TÉCNICA DE DATOS DE PANEL 8

ALTERNATIVAS DE ESPECIFICACIÓN DE DATOS DE PANEL A PARTIR DEL MODELO GENERAL. **9**

MODELO DE EFECTOS FIJOS 11

MODELO DE EFECTOS ALEATORIOS 11

ELECCIÓN DEL MÉTODO: ¿EFECTOS FIJOS O EFECTOS ALEATORIOS? 12

APLICACIÓN **13**

UNA FUNCIÓN DE PRODUCCIÓN PARA PAÍSES DEL MERCOSUR CON SOFTWARE E-VIEWS 6.0 13

SOLUCIÓN *13*

BIBLIOGRAFÍA **24**

Introducción

En el análisis de la información (económica, social, empresarial, comercial, etc.) pueden existir diferentes dimensiones sobre las cuales interesa obtener conclusiones derivadas de la estimación de modelos que traten de extraer relaciones de causalidad o de comportamiento entre diferentes tipos de variables, a partir de los datos disponibles.

Una de estas dimensiones la constituye el análisis de series de tiempo, la cual incorpora información de variables y/o unidades individuales de estudio durante un período determinado de tiempo (dimensión temporal). En este caso, cada período de tiempo constituye el elemento poblacional y/o muestral. Ejemplo de este tipo de datos lo constituyen las series del PIB y de tasas de interés de un país o el número de llamadas telefónicas de una familia a lo largo de un período determinado de tiempo.

Especificación de la relación:

$$Y_t = a + b_1 X_{1t} + b_2 X_{2t} + \cdots + b_K X_{kt} + u_t; \qquad \text{con} \quad t = 1, \dots, T$$

Tabla de datos:

$U.O$	Y	X_1	X_2	$\cdots$	X_k
1			X_{21}		
2			X_{22}		
$\vdots$					
t			X_{2t}		
$\vdots$					
T			X_{2T}		

Supuestos a verificar:

Sobre la perturbación aleatoria

- Media nula
- Homocedasticidad
- No autocorrelación
- Normalidad

Sobre la parte sistemática de la relación

- Linealidad
- No multicolinealidad
- Regresores no estocásticos

Existe otra dimensión que no incorpora el aspecto temporal sino que más bien representa el análisis de la información para las unidades individuales de estudio, en un momento determinado del tiempo (dimensión estructural). En este tipo de análisis, el cual se denomina de corte transversal, el elemento o unidad muestral no lo constituye el tiempo sino las unidades de análisis. Ejemplos de este tipo de análisis pueden ser la cantidad demandada de alimentos por una muestra de familias durante un período de tiempo, o la cantidad producida de televisores por una serie de empresas en el mismo lapso. En ambos casos los elementos muestrales son la familia y la empresa.

Especificación de la relación:

$$Y_i = a + b_1X_{1i} + b_2X_{2i} + \cdots + b_KX_{ki} + u_i; \qquad \text{con} \quad i = 1, \dots, n$$

Tabla de datos:

$U.O$	Y	X_1	X_2	$\cdots$	X_k
1			X_{21}		
2			X_{22}		
$\vdots$					
i			X_{2i}		
$\vdots$					
n			X_{2n}		

Supuestos a verificar:

Sobre la perturbación aleatoria

- Media nula
- Homocedasticidad
- Normalidad

Sobre la parte sistemática de la relación

- Linealidad
- No multicolinealidad
- Regresores no estocáticos

Ambos tipos de análisis de la información permiten extraer conclusiones relevantes de acuerdo con los intereses del investigador. Un sencillo ejemplo puede ilustrar mejor la diferencia entre cada uno de los enfoques: supóngase que se quiere modelar la Producción del Sector Agrícola de los países que integran el Mercosur.

Un análisis de regresión basado en datos de corte transversal para un año en particular, por ejemplo año 2009, puede incluir una serie de variables explicativas tales como el capital utilizado, la tierra afectada a la explotación agrícola y la mano de obra ocupada en dicha explotación; todas medidas en unidades físicas.

Con el tipo de información incluida en este modelo de corte transversal se podría estar tomando en cuenta cualquier tipo de economía de escala de la que los países podrían beneficiarse. Sin embargo, este modelo no puede identificar o tomar en cuenta como variable explicativa de la producción, cualquier incremento en el rendimiento que pueda ocurrir con el transcurso del tiempo como consecuencia de mejoras tecnológicas que hayan sido incorporadas en la industria.

La técnica de datos de panel

A diferencia de lo mencionado en el aparte anterior, un modelo econométrico de datos de panel es uno que incluye una muestra de agentes económicos o de interés (individuos, empresas, bancos, ciudades, países, etc.) para un período determinado de tiempo, esto es, combina ambos tipos de datos (dimensión temporal y estructural).

A manera de ejemplo, se puede disponer de datos anuales de producción agrícola de Argentina, Brasil, Paraguay y Uruguay durante un período de 20 años (1990 – 2009), lo cual es una base de datos mixta de serie temporal y corte transversal constituyéndose en un panel de datos. En este ejemplo, los elementos muestrales son el tiempo y los países del Mercosur.

El principal objetivo de aplicar y estudiar los datos en panel, es capturar la heterogeneidad no observable, ya sea entre agentes económicos o de estudio así como también en el tiempo, dado que esta heterogeneidad no se puede detectar ni con estudios de series temporales ni tampoco con los de corte transversal.

Esta técnica permite realizar un análisis más dinámico al incorporar la dimensión temporal de los datos, lo que enriquece el estudio, particularmente en períodos de

grandes cambios. Esta modalidad de analizar la información en un modelo de panel es muy usual en estudios de naturaleza microeconómica.

La aplicación de esta metodología permite analizar dos aspectos de suma importancia cuando se trabaja con este tipo de información y que forman parte de la heterogeneidad no observable: i) los efectos individuales específicos y ii) los efectos temporales.

En lo que se refiere a los efectos individuales específicos, se dice que estos son aquellos que afectan de manera desigual a cada uno de los agentes de estudio contenidos en la muestra (individuos, empresas, países) los cuales son invariables en el tiempo y que afectan de manera directa las decisiones que tomen dichas unidades. Usualmente se identifica este tipo de efectos con cuestiones de capacidad empresarial, eficiencia operativa, capitalización de la experiencia, acceso a la tecnología, etc.

Los efectos temporales son aquellos que afectan por igual a todas las unidades individuales del estudio. Este tipo de efectos pueden asociarse, por ejemplo, a los shocks macroeconómicos que pueden afectar por igual a todas las empresas o unidades de estudio.

Especificación general de un modelo de datos de panel.

La especificación general de un modelo de regresión con datos de panel es la siguiente:

$$Y_{it} = a_{it} + b_1 X_{1it} + b_2 X_{2it} + \cdots b_K X_{kit} + U_{it}; \qquad \text{con } i = 1, \dots, n \ y \ \ t = 1, \dots, T \qquad (1)$$

Donde i se refiere al individuo o a la unidad de estudio (corte transversal), t a la dimensión en el tiempo,

a es un vector de interceptos que puede contener entre 1 y $n + t$ parámetros,

b es un vector de K parámetros y

X_{it} es la i-ésima observación al momento t para las K variables explicativas $X_1, X_2, \dots, X_K$.

En este caso, la muestra total de las observaciones en el modelo viene dado por $n \ x \ T$.

A partir de este modelo general, y con base en ciertos supuestos y restricciones acerca del valor de algunos de los parámetros, se pueden derivar algunas otras variantes de modelos de datos de panel, las cuales se describirán con más detalle en una sección posterior.

Tabla de datos:

$Tiempo$	$País$	Y	X_1	X_2	$\cdots$	X_k
1	ARG			X_{2ARG1}		
2	ARG			X_{2ARG2}		
$\vdots$						
t	ARG			X_{2ARGt}		
1	BRA			X_{2BRA1}		
2	BRA			X_{2BRA2}		
$\vdots$						
t	BRA			X_{2BRAt}		
1	URU			X_{2URU1}		
2	URU			X_{2URU2}		
$\vdots$						
t	URU			X_{2URUt}		
1	PAR			X_{2PAR1}		
2	PAR			X_{2PAR2}		
$\vdots$						
t	PAR			X_{2PARt}		

a) Componentes del error

Es usual interpretar los modelos de datos de panel a través de sus componentes de errores. El término de error U_{it} incluido en la ecuación (1), puede descomponerse de la siguiente manera:

$$U_{it} = u_i + v_t + w_{it} \qquad (2)$$

u_i representa los efectos no observables que difieren entre las unidades de estudio pero no en el tiempo, que generalmente se los asocia a la tecnología incorporada.

v_t se le identifica con efectos no cuantificables que varían en el tiempo pero no entre las unidades de estudio.

w_{it} se refiere al término de error puramente aleatorio.

La mayoría de las aplicaciones con datos de panel utilizan el modelo de componente de error conocido como "one way" para el cual $v_t = 0$. Este tipo de análisis supone que no existen efectos no cuantificables que varíen en el tiempo pero no entre las unidades individuales de estudio. Existe además el modelo "two-way" en el cual el componente de error $v_t \neq 0$, a través del cual se pretende capturar efectos temporales específicos (choques) que no están incluidos en la regresión.

Las diferentes variantes para el modelo "one way" de componentes de errores surgen de los distintos supuestos que se hacen acerca del término u_i. Pueden presentarse tres posibilidades:

- El caso más sencillo es el que considera al $u_i = 0$, o sea, no existe heterogeneidad no observable entre los individuos o firmas. Dado lo anterior, los U_{it} satisfacen todos los supuestos del modelo lineal general, por lo cual el método de estimación de mínimos cuadrados clásicos produce los mejores estimadores lineales e insesgados.
- La segunda posibilidad consiste en suponer a u_i un efecto fijo y distinto para cada unidad de corte transversal. En este caso, la heterogeneidad no observable se incorpora a la constante del modelo.
- La tercera alternativa es tratar a u_i como una variable aleatoria no observable que varía entre individuos pero no en el tiempo.

b) Ventajas y Desventajas de la Técnica de Datos de Panel

La técnica de datos de panel presenta una serie de ventajas y desventajas en comparación con los modelos de series de tiempo y de corte transversal. Las más relevantes son las siguientes:

Ventajas:

- La técnica permite al investigador económico disponer de un mayor número de observaciones $(n\ x\ T)$ incrementando los grados de libertad, reduciendo la colinealidad entre las variables explicativas y, en última instancia, mejorando la eficiencia de las estimaciones econométricas.
- Tal y como se mencionó anteriormente, la técnica permite capturar la heterogeneidad no observable ya sea entre unidades individuales de estudio como en el tiempo. Con base en lo anterior, la técnica permite aplicar una serie de pruebas de hipótesis para confirmar o rechazar dicha heterogeneidad y cómo capturarla.

- Los datos en panel suponen e incorporan, en el análisis, el hecho de que los individuos, firmas, bancos o países son heterogéneos. Los análisis de series de tiempo y de corte transversal no tratan de controlar esta heterogeneidad corriendo el riesgo de obtener resultados sesgados.
- Permite estudiar de una mejor manera la dinámica de los procesos de ajuste. Esto es fundamentalmente cierto en estudios sobre el grado de duración y permanencia de ciertos niveles de condición económica (desempleo, pobreza, riqueza).
- Permite elaborar y probar modelos relativamente complejos de comportamiento en comparación con los análisis de series de tiempo y de corte transversal. Un ejemplo claro de este tipo de modelos, son los que se refieren a los que tratan de medir niveles de eficiencia técnica por parte de unidades económicas individuales (empresas, bancos, etc.).

Desventajas:

En términos generales, las desventajas asociadas a la técnica de datos de panel se relacionan con los procesos para la obtención y el procesamiento de la información estadística sobre las unidades individuales de estudio, cuando esta se obtiene por medio de encuestas, entrevistas o utilizando algún otro medio de relevamiento de los datos. Ejemplos de este tipo de limitaciones son: cobertura de la población de interés, porcentajes de respuesta, preguntas confusas, distorsión deliberada de las respuestas, etc.

Alternativas de especificación de datos de panel a partir del modelo general.

La primera especificación (A) se refiere al caso en que no existe heterogeneidad no observable en el sistema de datos de panel y por tanto se emplea el método de Mínimos Cuadrados Ordinarios con la ventaja de ganar grados de libertad.

Ahora bien, en los casos en que se rechaza la hipótesis de homogeneidad en un sistema de datos de panel, es decir que existe heterogeneidad no observable ya sea a través del tiempo, entre unidades de estudio (individuos) o en ambos sentidos, debe buscarse una especificación que la capture en forma apropiada, con el fin de evitar el problema de sesgo sobre los estimadores de los parámetros de las variables explicativas, que se cometería si se emplea la especificación.

Existen dos procedimientos adicionales para estimar el modelo en un sistema de datos de panel: uno de ellos implica el reconocimiento de que las variables omitidas pueden generar cambios en los interceptos ya sea a través del tiempo o entre unidades de corte transversal, en este caso el modelo de efectos fijos trata de aproximar estos

cambios con variables dummy; el otro modelo es el de efectos aleatorios, que trata de capturar estas diferencias a través del componente aleatorio del modelo.

Como ya se mencionó, la técnica de datos de panel permite contemplar la existencia de efectos individuales específicos a cada unidad de corte transversal, invariables en el tiempo que afectan la manera en que cada unidad de corte transversal toma sus decisiones.

Una forma simple, y de hecho la más utilizada, de considerar esta heterogeneidad es empleando los modelos de intercepto variable. Así, el modelo lineal es el mismo para todas las unidades o individuos bajo estudio, pero la ordenada al origen es específica a cada una de ellas. A partir del modelo general esta situación se representa mediante la siguiente ecuación:

$$Y_{it} = a_i + b_1 X_{1it} + b_2 X_{2it} + \cdots b_K X_{kit} + U_{it}; \qquad \text{con } i = 1, \dots, n \ y \ \ t = 1, \dots, T \qquad (3)$$

El supuesto básico de estos modelos es que dadas las variables explicativas observadas, los efectos de todas las variables omitidas pueden representarse de tres formas posibles:

a. Una variable por cada individuo, no variable en el tiempo: este es el caso de variables que son las mismas para cada unidad de corte transversal a través del tiempo.
b. Una variable por periodo pero no variables entre individuos: son la misma variable para todos los individuos en un momento del tiempo pero varían a lo largo del periodo de estudio.
c. Una variable que cambia en el tiempo y por individuo: se trata de variables que cambian entre individuos en un momento del tiempo, y que además cambian a lo largo del tiempo.

Estos modelos de intercepto variable asumen que los efectos de las variables omitidas, ya sean específicas a cada individuo y/o que cambian en el tiempo, no son importantes en forma individual, pero que si son importantes si se consideran en conjunto.

Por otro lado, dado que el efecto de las variables omitidas puede mantenerse constante en el tiempo para cada individuo, o ser el mismo para todos los individuos en un momento en el tiempo, o una combinación de ambos, se pueden capturar en el término constante de un modelo de regresión como un promedio que toma en cuenta explícitamente la heterogeneidad entre individuos y/o en el tiempo contenida en los datos (este es el procedimiento que se emplea por ejemplo al estimar una función Cobb Douglas).

Seguidamente, se analizarán los principales modelos que se estudian a partir de la especificación general y de acuerdo con la forma de incorporar la heterogeneidad no observada.

Modelo de efectos fijos

Como se indicó brevemente, una posibilidad es explicar los datos con el modelo de efectos fijos el cual considera que existe un término constante diferente para cada individuo y supone que los efectos individuales son independientes entre sí.

Con este modelo se considera que las variables explicativas afectan por igual a las unidades de corte transversal y que éstas se diferencian por características propias de cada una de ellas, medidas por medio del intercepto. Es por ello que los n interceptos se asocian con variables dummy con coeficientes específicos para cada unidad, los cuales se deben estimar. Para la i-ésima unidad de corte transversal, la relación es la siguiente:

$$Y_i = ia_i + b_1X_{1i} + b_2X_{2i} + \cdots b_KX_{ki} + U_i; \qquad \text{con } i = 1, \dots, n \qquad (4)$$

Donde el subíndice i representa un vector columna de unos. Debe hacerse notar que en este modelo se presenta una pérdida importante de grados de libertad.

Modelo de efectos aleatorios

A diferencia del modelo de efectos fijos, el modelo de efectos aleatorios considera que los efectos individuales no son independientes entre sí, sino que están distribuidos aleatoriamente alrededor de un valor dado. Una práctica común en el análisis de regresión es asumir que el gran número de factores que afecta el valor de las variable dependiente pero que no han sido incluidas explícitamente como variables independientes del modelo, pueden resumirse apropiadamente en la perturbación aleatoria.

Así, con este modelo se considera que tanto el impacto de las variables explicativas como las características propias de cada unidad de corte transversal son diferentes. El modelo se expresa algebraicamente de la siguiente forma:

$$Y_{it} = (a + u_i) + b_1X_{1it} + b_2X_{2it} + \cdots b_KX_{kit} + w_{it}; \qquad \text{con } i = 1, \dots, n \quad t = 1, \dots, T \qquad (5)$$

Donde: u_i viene a representar la perturbación aleatoria que permitiría distinguir el efecto de cada individuo en el panel. Para efectos de su estimación se agrupan los componentes estocásticos, y se obtiene la siguiente relación:

$$Y_{it} = a + b_1X_{1it} + b_2X_{2it} + \cdots b_KX_{kit} + U_{it}; \qquad \text{con } i = 1, \dots, n \quad t = 1, \dots, T \qquad (6)$$

Donde $U_{it} = u_i + v_t + w_{it}$ se convierte en el nuevo término de la perturbación, U no es homocedástico, donde $u_i,\ v_t,\ w_{it}$ corresponden al error asociado con las series de tiempo (v_t); a la perturbación de corte transversal (u_i) y el efecto aleatorio combinado de ambas (w_{it}).

El método de Mínimos Cuadrados Ordinarios (MCO) no es aplicable dado que no se cumplen los supuestos que permiten que el estimador sea consistente. Por lo que es preferible en este caso utilizar el método de Mínimos cuadrados Generalizados (MCG) cuyas estimaciones son superiores al de MCO en caso de no cumplirse los supuestos tradicionales y son similares en caso contrario.

Elección del método: ¿efectos fijos o efectos aleatorios?

La decisión acerca de la estructura apropiada para el análisis, es decir, efectos fijos versus efectos aleatorios depende en parte de los siguientes aspectos:

a) Los objetivos del estudio

Si se desea hacer inferencias con respecto a la población, es decir que se trabaja con una muestra aleatoria, lo mejor es utilizar una especificación del tipo aleatoria. En caso de que el interés sea limitado a una muestra que se ha seleccionado a conveniencia o bien que se está trabajando con la población, la estimación de efectos fijos será la correcta.

Adicionalmente, si el interés del estudio particular está puesto en los coeficientes de las pendientes de los parámetros, y no tanto en las diferencias individuales, se debería elegir un método que relegue estas diferencias y tratar la heterogeneidad no observable como aleatoria.

El modelo de efectos fijos se ve como un caso en que el investigador hace inferencia condicionada a los efectos que ve en la muestra. El de efectos aleatorios se ve como uno en el cual el investigador hace inferencia condicional o marginal respecto a una población.

Se deja al investigador que decida si hace inferencia con respecto a las características de una población o solo respecto a los efectos que están en la muestra.

b) El contexto de los datos, es decir, cómo fueron obtenidos y el entorno de donde provienen

Con el método de efectos fijos la heterogeneidad no observable se incorpora en la ordenada al origen del modelo y con la de efectos aleatorios, como ya se mencionó, se incorporan en el término de error, por lo cual lo que se modifica es la varianza del modelo.

Emplear un modelo de efectos fijos o aleatorios genera diferencias en las estimaciones de los parámetros en los casos en que se cuenta con t pequeño y n grande. En estos

casos debe hacerse el uso más eficiente de la información para estimar esa parte de la relación de comportamiento contenida en las variables que difieren sustancialmente de un individuo a otro.

c) Número de datos disponibles

El método de efectos fijos presenta el problema de que el uso de variables "Dummies" no identifica directamente qué causa que la regresión lineal cambie en el tiempo y en los individuos. Además, esto implica la pérdida de grados de libertad.

Asimismo, deberán tomarse consideraciones con respecto a la estructura de los datos con que se cuente, dado que si la n es grande pero si se tiene un t pequeño, podría ser que el número de parámetros de efectos fijos sea muy grande en relación con el número de datos disponibles, con parámetros poco confiables y una estimación ineficiente.

Algunas investigaciones han demostrado que el emplear modelos de efectos fijos produce resultados significativamente diferentes al de efectos aleatorios cuando se estima una ecuación usando una muestra de muchas unidades de corte transversal con pocos periodos de tiempo (629 individuos para 6 periodos, por ejemplo).

APLICACIÓN

UNA FUNCIÓN DE PRODUCCIÓN PARA PAÍSES DEL MERCOSUR CON SOFTWARE E-VIEWS 6.0

La siguiente tabla de datos contiene la producción en toneladas de productos agrícolas (AGRIPRO), el capital asignado a la producción en unidades físicas (AGRICAP), el personal ocupado en número de personas (AGRITRAB) y la tierra utilizada para la producción en has sembradas (AGRITIERRA), para 4 países del Mercosur durante 20 años (1990 – 2009).

Con esta información se trata de explicar la producción en función de los factores capital, trabajo y tierra (todas las variables en logaritmos) a partir del panel de datos de los cuatro países (1. Argentina, 2. Brasil, 3 Paraguay, 4. Uruguay). A través de especificar distintos modelos se trata de encontrar el ajuste más adecuado para este conjunto de países en el periodo de tiempo considerado.

SOLUCIÓN

Se especifica el modelo de datos de panel siguiente:

$Log(AGRIPRO_{it}) = a_{it} + b_1 Log(AGRICAP_{it}) + b_2 Log(AGRITIERRA_{it}) + b_3 Log(AGRITRAB_{it}) + U_{it}; i = 1, \dots, 4 \quad t = 1990, \dots, 2009$

Años	País	AGRIPRO	AGRICAP	AGRITIERRA	AGRITRAB
1990	Argentina	27658895	52309	12736482	708232
1991	Argentina	30917778	54810	13376579	716429
1992	Argentina	33491430	50275	12483142	724725
1993	Argentina	33025328	50809	12488627	733121
1994	Argentina	33250356	51802	13178787	730769
1995	Argentina	37132403	53412	14184974	750074
1996	Argentina	34962758	54634	14640979	760103
1997	Argentina	44958835	59542	18049801	769935
1998	Argentina	55194648	58103	16724013	774007
1999	Argentina	48563335	57659	16378414	758242
2000	Argentina	53908107	60006	17133093	747917
2001	Argentina	59845128	61101	18361127	744909
2002	Argentina	61209293	55269	18764715	730523
2003	Argentina	63299207	57150	18379301	738099
2004	Argentina	62655691	57637	19143054	762963
2005	Argentina	76315977	58002	19582820	771822
2006	Argentina	69510334	59796	18675902	776884
2007	Argentina	85739810	62479	19769185	781272
2008	Argentina	86678915	64384	20559992	799940
2009	Argentina	54523032	69713	19418653	802485
1990	Brasil	56687998	157482	33888100	8442986
1991	Brasil	56224570	152278	31983825	9731144
1992	Brasil	67448010	150884	32940272	11019302
1993	Brasil	67709031	143472	30896774	10860439
1994	Brasil	75463078	150500	32786890	10840479
1995	Brasil	80272166	148560	32565353	10820519
1996	Brasil	70920728	139731	29930807	9925481
1997	Brasil	76610485	133161	29940046	9990009
1998	Brasil	74844347	127810	29293964	9747760
1999	Brasil	81860371	129323	30341768	10568391
2000	Brasil	82385097	129356	31208411	9333882
2001	Brasil	98844948	131816	31614909	9255361
2002	Brasil	96170204	138796	33314913	9689905
2003	Brasil	121319749	145864	36087121	10144851
2004	Brasil	117539782	154170	40182506	10558993
2005	Brasil	110698428	151081	41236779	10606607
2006	Brasil	115347538	147117	38660836	10279247
2007	Brasil	131681848	150663	37526462	10481616
2008	Brasil	144018276	156864	38523443	10455823
2009	Brasil	132668986	160798	38569000	10405562
1990	Paraguay	7511371	4952	2082	199448
1991	Paraguay	8948556	5085	1678	228049
1992	Paraguay	9022564	5255	1792	201453
1993	Paraguay	10005304	5423	1661	202707
1994	Paraguay	9657841	5593	1857	227939
1995	Paraguay	10799234	5764	1922	253247
1996	Paraguay	10602932	5938	2020	272550
1997	Paraguay	11710661	6118	1989	317510
1998	Paraguay	11845140	6315	2201	314473
1999	Paraguay	12358600	6510	2225	262151
2000	Paraguay	10473173	6616	2178	296432
2001	Paraguay	12463720	6653	2482	330261
2002	Paraguay	13467744	6952	2281	390539
2003	Paraguay	15429292	7838	2744	417945
2004	Paraguay	16361417	8688	3335	538333
2005	Paraguay	15090744	9060	3523	523481
2006	Paraguay	15266868	9336	3881	470605
2007	Paraguay	18428360	10008	3718	469660
2008	Paraguay	18533754	11261	3768	410056
2009	Paraguay	15768421	10962	4075	511656
1990	Uruguay	1370	82599	11659	584
1991	Uruguay	1272	90371	11778	527
1992	Uruguay	1695	85046	11899	588
1993	Uruguay	1367	83599	12021	589

continúa…

					Continuación
1994	Uruguay	1863	90198	12144	622
1995	Uruguay	1745	92391	12269	642
1996	Uruguay	2433	91075	12395	756
1997	Uruguay	2162	92786	12523	716
1998	Uruguay	2779	94262	12667	711
1999	Uruguay	1966	97548	12798	557
2000	Uruguay	2226	101006	12651	530
2001	Uruguay	1819	100209	12611	623
2002	Uruguay	2132	99972	12695	705
2003	Uruguay	2875	99175	12876	845
2004	Uruguay	3272	98135	13422	979
2005	Uruguay	3128	97393	13599	841
2006	Uruguay	3658	96651	13744	973
2007	Uruguay	3946	95605	14309	1171
2008	Uruguay	5059	95659	14796	1561
2009	Uruguay	6194	94916	14811	1882

Para la estimación se trabaja con el software $E - Views$, el cual se abre haciendo doble click en el ícono existente en el escritorio o desde *Inicio-Todos los programas-Eviews 6-Eviews6*. Dentro de *Eviews*, el primer paso es incorporar al software la tabla de datos de $Excel$ existente en el archivo *agrícola.xls*. El formato del archivo de Excel que contiene los datos debe corresponder a la versión 97-2003; es necesario tomar nota de la celda donde se encuentra el primer dato (C2), de la cantidad de columnas con variables de interés a importar (6), el tipo de dato (anual desde 1990), la hoja en la cual está la información (Hoja 1).

En *Eviews* ir a *File-Open-Foreign data as Workfile…* (Figura 1), se despliega el explorador de windows para localizar el archivo; luego se debe seleccionar el archivo agrícola.xls y *Abrir.* Esta acción despliega la pantalla de lectura de la de Cálculo para convertir desde lenguaje *Excel* a lenguaje *Eviews* (Figura 2). En esta pantalla es posible indicarle la hoja donde se encuentra la información en el archivo Excel y observar si está leyendo las columnas que es necesario importar. En la pantalla siguiente hay que aceptar la configuración por defecto que ofrece el software desde el ícono *Finalizar* (Figura 3). El resultado es disponer de los datos en el entorno *Eviews* (Figura 4); al desplazar hacia abajo la barra del lado derecho de la ventana del grupo de datos se observa que, a continuación de los datos de Argentina, se encuentran los datos de Brasil, Paraguay y Uruguay. En la ventana *Workfile* se encuentran los íconos de cada serie de datos (las variables de análisis AGRICAP AGRIPRO, AGRITIERRA y AGRITRAB), en el encabezado de la ventana se informa que el archivo tiene 80 observaciones que corresponden al total de 4 países en 20 años; esto indica que el archivo no tiene formato de panel.

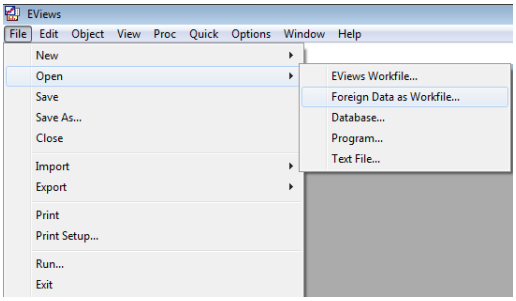


Figura 1

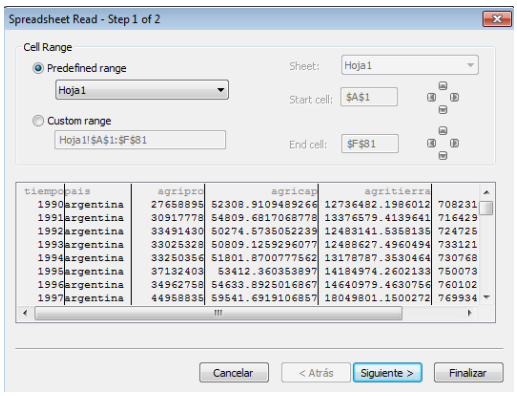


Figura 2

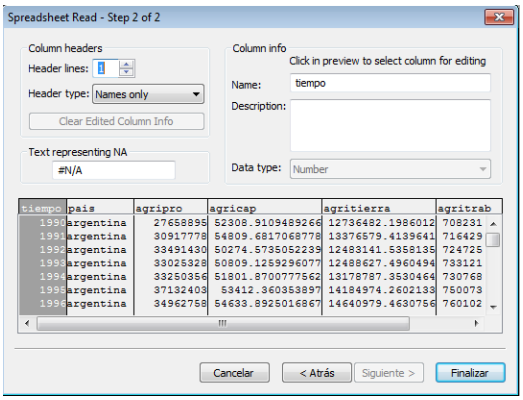


Figura 3

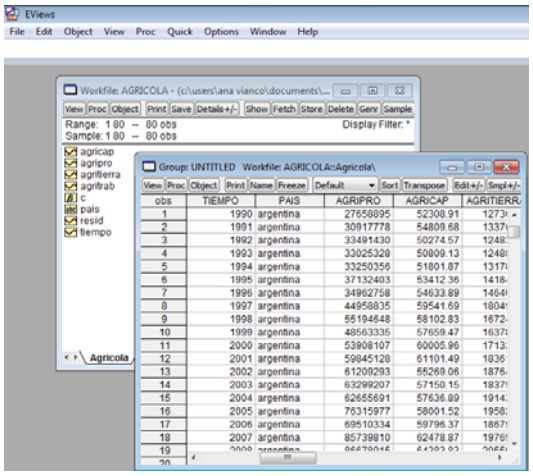


Figura 4

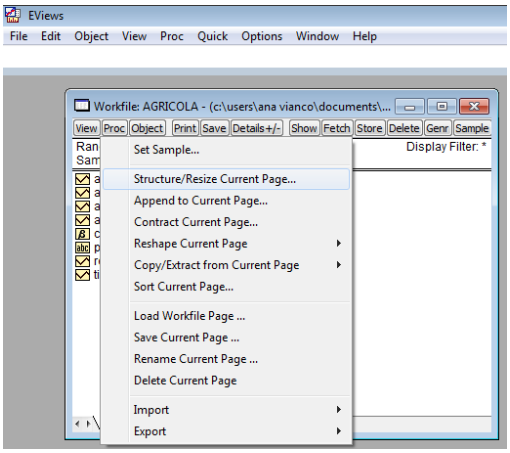


Figura 5

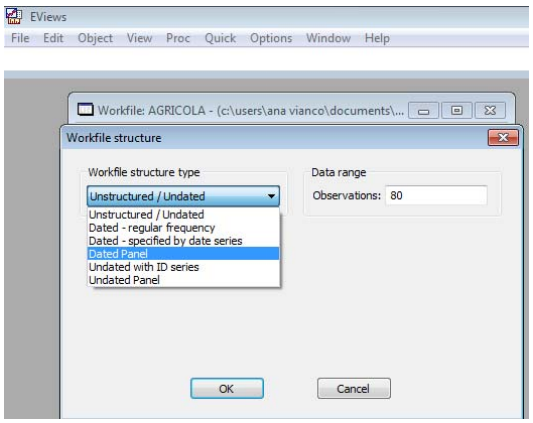


Figura 6

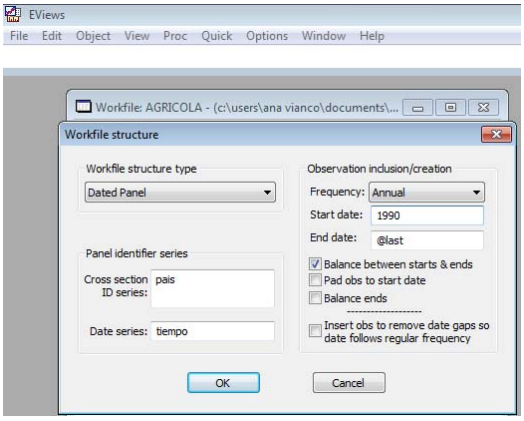


Figura 7

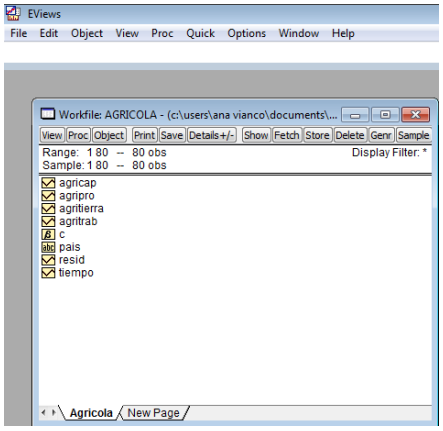


Figura 8

Para dar formato de panel hay que trabajar en la ventana *Workfile*; ir a *Procs-Structure/Resize Current Page* (Figura 5), se muestra la ventana *Workfile structure* y al desplegar las alternativas de *Workfile structure type* se selecciona *Dated Panel* (Figura 6); en la nueva pantalla se debe indicar la frecuencia anual con el primer año de la serie (1990) y las variables que identifican la

estructura del panel: *Cross section – país* y *Date series – tiempo* (Figura 7). Estas acciones cambian el encabezado de la ventana *Workfile: Range 1990 2009 x 4 – 80 obs* (Figura 8). Para grabar este entorno de trabajo, ir a *File-Save As*, seleccionar la carpeta donde guardar el archivo y dar un nombre. Al grabar, *Eviews* genera sólo dos archivos con igual nombre y distinta extensión; el archivo ejecutable es **.wf1*.

La tarea de análisis comienza con las representaciones gráficas de las variables del panel mediante $Quick \rightarrow Graph \rightarrow Line\ Graph$ (Figura 9) e introduciendo la variable Log(AGRIPRO) en la pantalla $Series\ List$ (Figura 10). Al pulsar OK se obtiene la pantalla $Graph\ Options$ (Figura 11); si se elige $Individual\ cross\ section\ data$ se obtiene la gráfica de Log(AGRIPRO) para todas las secciones cruzadas (Figura 12). En la pantalla $Graph - Options$, al elegir $Mean\ plus\ SD\ Bounds$ (Figura 13) se obtiene un gráfico de la evolución de las medias de la variable Log(AGRIPRO), en las distintas secciones cruzadas, con una banda de confianza de $\pm 2(\ desviaciones\ típicas)$ según se ve en la Figura 14.

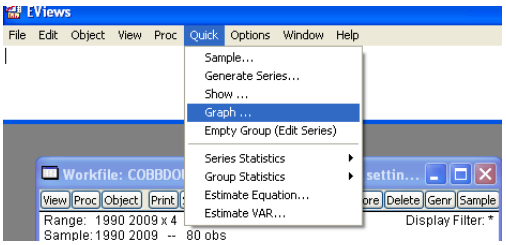


Figura 9

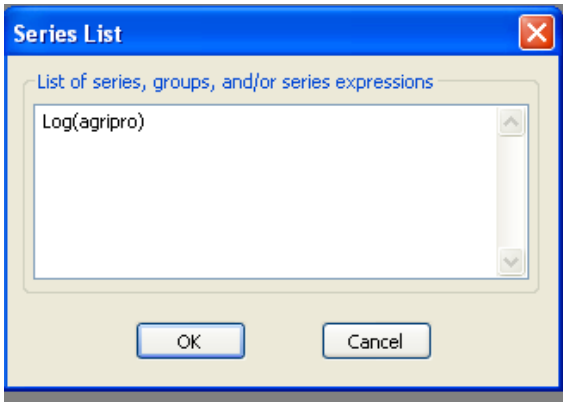


Figura 10

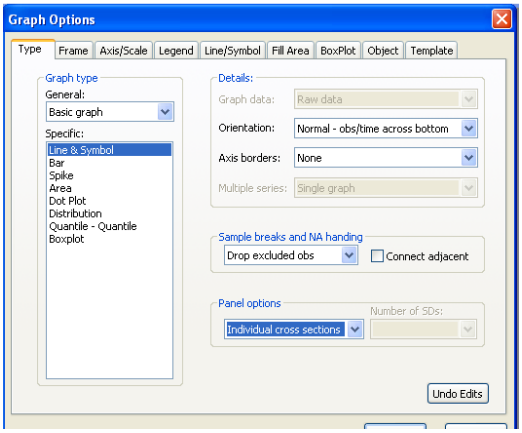


Figura 11

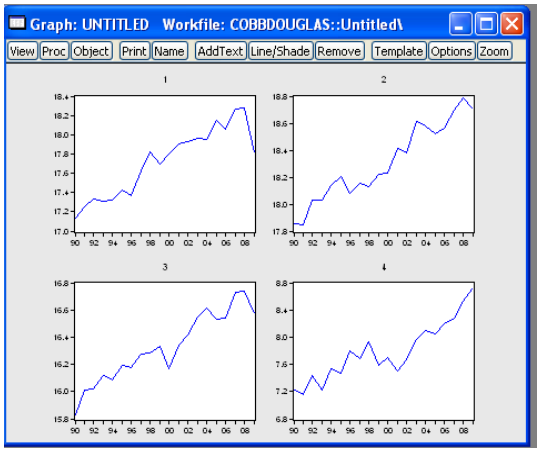


Figura 12

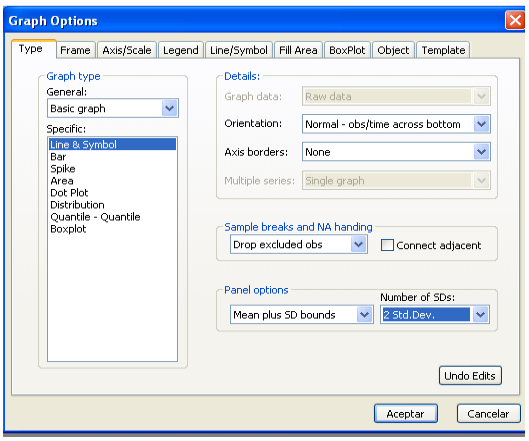


Figura 13

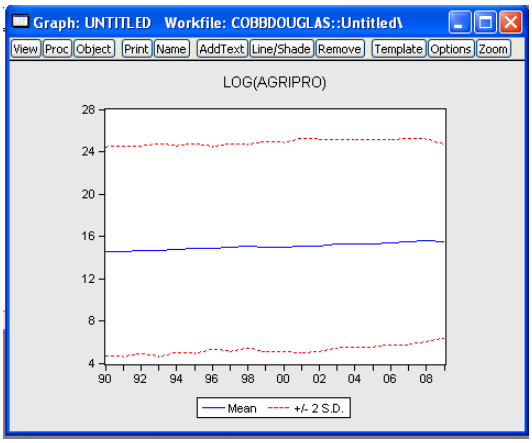


Figura 14

Para hacer contrastes de igualdad de medias o varianzas de Log(AGRIPRO), con doble click en AGRIPRO aparece la serie para los distintos países (Figura 15), luego basta seleccionar $View \rightarrow Descriptive\ Stats\ \&\ Tests \rightarrow Stats\ by\ Classification$ (Figura 16).

En la pantalla $Tests\ By\ Classification$ se elige la variable cuya igualdad de medias, medianas o varianzas en las distintas secciones cruzadas del panel se contrasta. En la Figura 17 se ha elegido contrastar medias y en la Figura 18 se ha elegido contrastar varianzas para Log(AGRIPRO). Las Figuras 19 y 20 muestran los resultados, observándose que se rechaza la igualdad de medias y la igualdad de varianzas (p-valores menores que 0,05). Lo mismo se puede repetir para todas las variables.

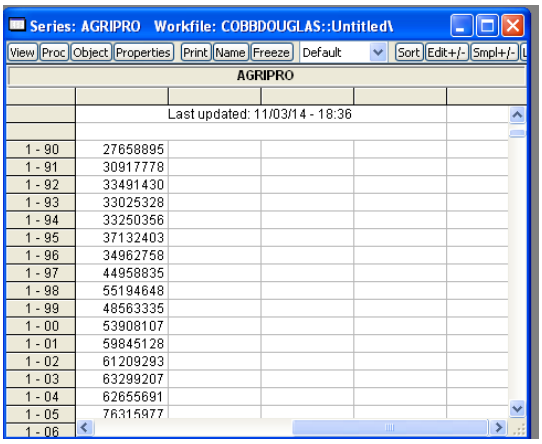


Figura 15

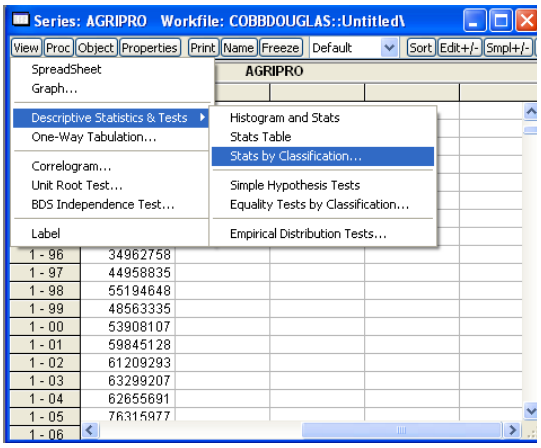


Figura 16

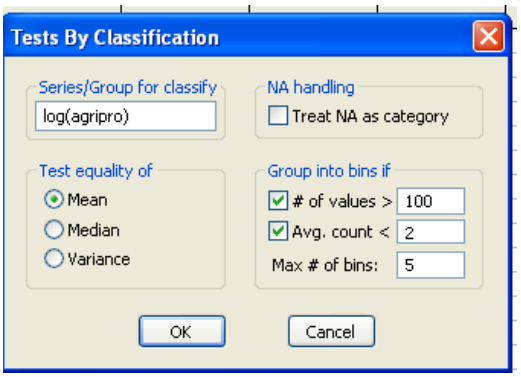


Figura 17

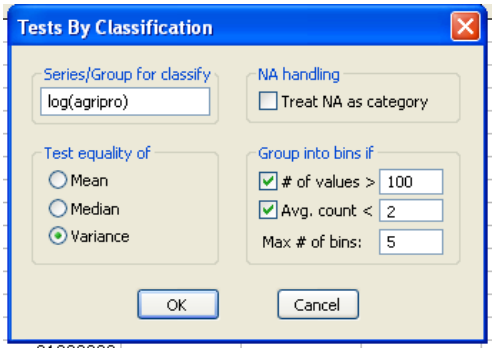
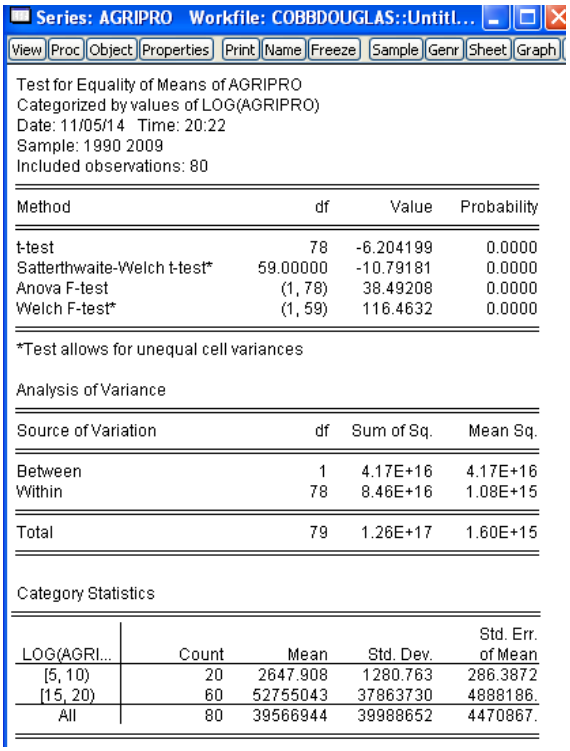


Figura 18



Series: AGRIPRO Workfile: COBBDOUGLAS::Untitl...

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph

Test for Equality of Means of AGRIPRO
Categorized by values of LOG(AGRIPRO)
Date: 11/05/14 Time: 20:22
Sample: 1990 2009
Included observations: 80

Method	df	Value	Probability
t-test	78	-6.204199	0.0000
Satterthwaite-Welch t-test*	59.00000	-10.79181	0.0000
Anova F-test	(1, 78)	38.49208	0.0000
Welch F-test*	(1, 59)	116.4632	0.0000

*Test allows for unequal cell variances

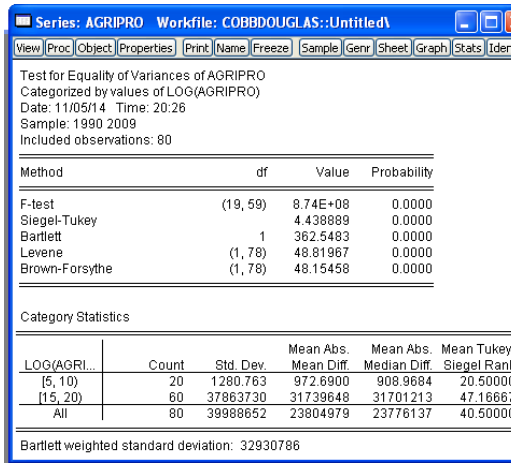
Analysis of Variance

Source of Variation	df	Sum of Sq.	Mean Sq.
Between	1	4.17E+16	4.17E+16
Within	78	8.46E+16	1.08E+15
Total	79	1.26E+17	1.60E+15

Category Statistics

LOG(AGRI...	Count	Mean	Std. Dev.	Std. Err. of Mean
[5, 10)	20	2647.908	1280.763	286.3872
[15, 20)	60	52755043	37863730	4888186.
All	80	39566944	39988652	4470867.

Figura 19



Series: AGRIPRO Workfile: COBBDOUGLAS::Untitled\

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph | Stats | Ider

Test for Equality of Variances of AGRIPRO
Categorized by values of LOG(AGRIPRO)
Date: 11/05/14 Time: 20:26
Sample: 1990 2009
Included observations: 80

Method	df	Value	Probability
F-test	(19, 59)	8.74E+08	0.0000
Siegel-Tukey		4.438889	0.0000
Bartlett	1	362.5483	0.0000
Levene	(1, 78)	48.81967	0.0000
Brown-Forsythe	(1, 78)	48.15458	0.0000

Category Statistics

LOG(AGRI...	Count	Std. Dev.	Mean Abs. Mean Diff.	Mean Abs. Median Diff.	Mean Tukey Siegel Ranl
[5, 10)	20	1280.763	972.6900	908.9684	20.5000
[15, 20)	60	37863730	31739648	31701213	47.1666
All	80	39988652	23804979	23776137	40.5000

Bartlett weighted standard deviation: 32930786

Figura 20

Una vez realizado el análisis gráfico y descriptivo de las variables del panel se procede a su estimación considerándolo inicialmente como un panel de coeficientes constantes. Para ello se elige $Quick \rightarrow Estimate\ Equation$, se escribe la ecuación del modelo a ajustar en el campo $Equation\ Specification$ de la solapa $Specification$, se elige $Least\ Squares$ en el campo $Method$ para ajustar por *mínimos cuadrados* (Figura 21), se rellena la solapa $Panel\ Options$ como se indica en la Figura 22 (sin especificación de efectos ni ponderaciones) y se hace clic en $Aceptar$. Se obtienen los resultados de la Figura 23.

Los resultados del ajuste del panel de la Figura 23 muestran una significatividad individual y conjunta de los coeficientes estimados muy alta y un coeficiente de determinación muy bueno. El mayor problema es el estadístico de Durbin Watson, que es muy bajo.

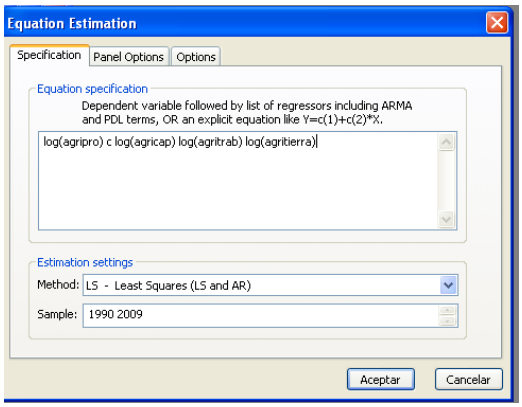


Figura 21

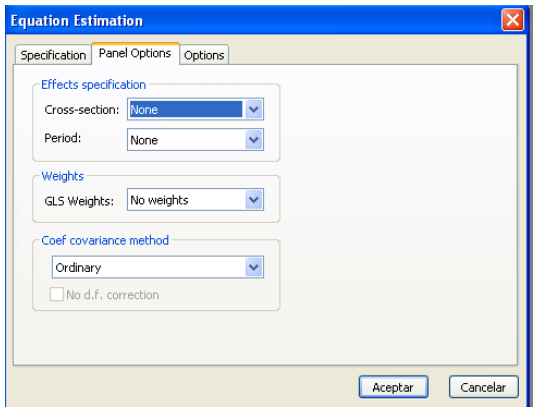


Figura 22

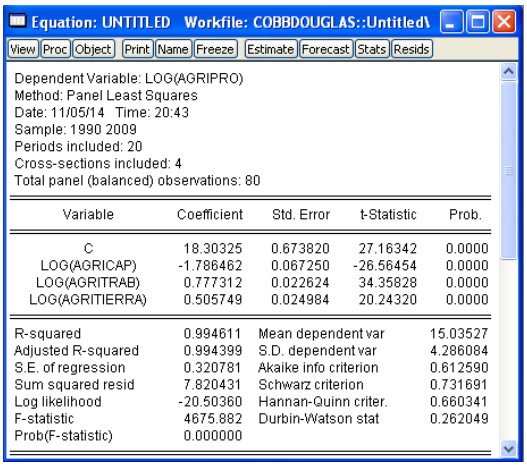


Figura 23

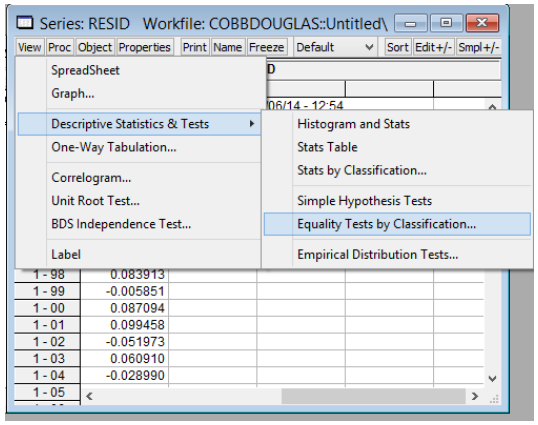


Figura 24

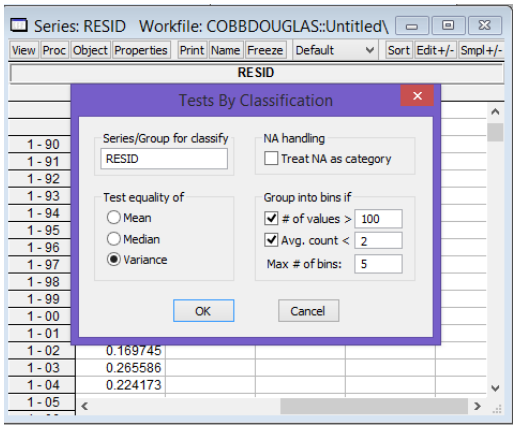


Figura 25

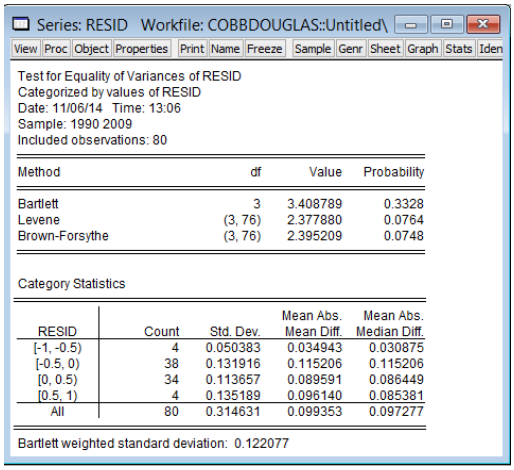


Figura 26

Para probar si hay *heteroscedasticiad* entre secciones cruzadas, hacer doble click sobre la variable RESID y elegir $View \rightarrow Descriptive\ Stats\ \&\ Tests \rightarrow Equality\ Tests\ by\ Classification$ (Figura 24). En la pantalla $Tests\ By\ Classification$ (Figura 25) se elige la variable cuya igualdad de varianzas en las distintas secciones cruzadas del panel se contrasta RESID. La Figura 26 muestra los resultados, observándose que se acepta la igualdad de varianzas residual en las distintas

secciones cruzadas (p-valor mayor que 0,05*). No existe entonces heteroscedasticidad entre secciones cruzadas.*

El panel de coeficientes constantes estimado es el siguiente:

$$Log(AGRIPRO)_{it} = 18{,}30325 - 1{,}786462 * \mathrm{Log}(AGRICAP)_{it} + 0{,}777312 * Log(AGRITRAB)_{it} + 0{,}505749 * Log(AGRITIERRA)_{it}$$

El problema de la *autocorrelación* puede intentar resolverse con la introducción de una estructura AR(1) en los residuos, haciendo clic en $Estimate$ y rellenado el campo $Equation\ Specification$ de la solapa $Specification$ como se indica en la Figura 27. Al hacer clic en $Aceptar$ se obtienen los resultados de la Figura 28 donde se observa un estadístico de Durbin Watson muy próximo a 2, buena significatividad conjunta de las variables, buen R2, pero se ha perdido significatividad individual.

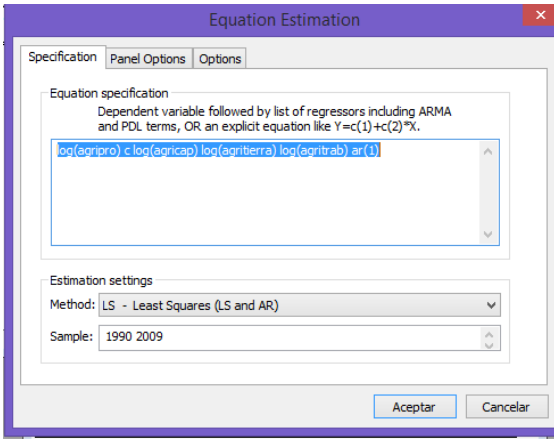


Figura 27

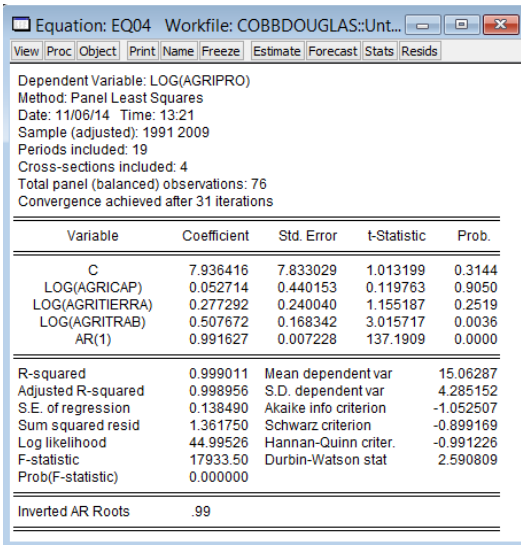


Figura 28

El panel de coeficientes constantes estimado sin autocorrelación sería el siguiente:

$$Log(AGRIPRO)_{it} = 7{,}936416 - 0{,}052714 * \mathrm{Log}(AGRICAP)_{it} + 0{,}507672 * Log(AGRITRAB)_{it} + 0{,}277992 * Log(AGRITIERRA)_{it} + 0{,}991627 * U_{it-1} + \varepsilon_{it}$$

Ahora se prueba con el modelo general de efectos fijos. Este modelo presenta los residuos

$$U_{it} = u_i + v_t + w_{it}$$

Siendo w_{it}, variable aleatoria y u_i (efectos de sección cruzada) y v_t (efectos de tiempo) deterministas.

Se estima el panel con efectos fijos de secciones cruzadas (efectos de país) y efectos fijos de tiempo. Para ello se elige $Quick \rightarrow Estimate\ Equation$, se escribe la ecuación del modelo a ajustar en el campo $Equation\ Specification$ de la solapa $Specification$ se elige $Least\ Squares$ en el campo $Method$ para ajustar por mínimos cuadrados (Figura 29), se rellena la solapa $Panel\ Options$ como se indica en la Figura 30 (se especifican efectos fijos de sección cruzada y de tiempo) y se hace clic en $Aceptar$.

Se obtienen los resultados de la Figura 31, donde se ve un estadístico de Durbin Watson razonable, buena significatividad individual y conjunta y buen R^2.

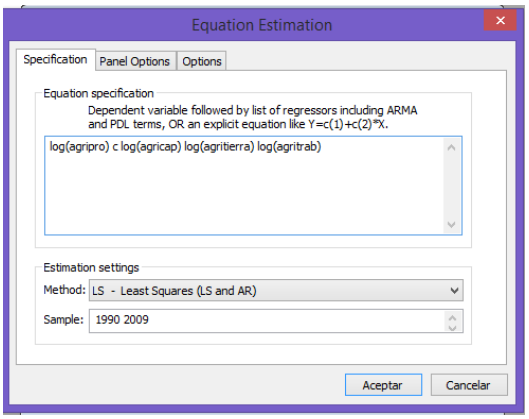


Figura 29

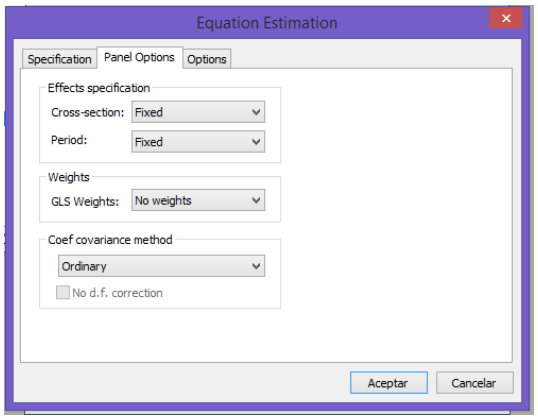


Figura 30

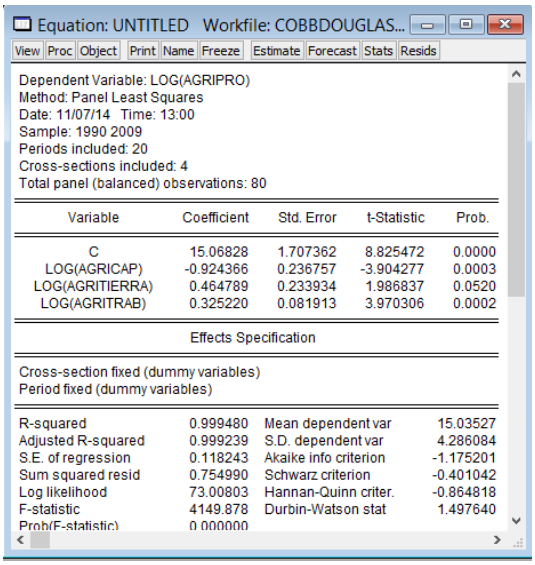


Figura 31

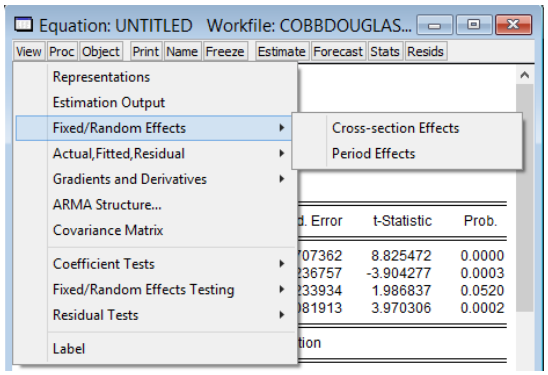


Figura 32

Si se quiere ver las estimaciones de los efectos fijos, por país y por unidades de tiempo (Figura 32), basta con elegir $View \rightarrow\ Fixed/Random\ Effects \rightarrow$ y en la Figura 33 y Figura 34, se observan dichos efectos, respectivamente.

Para probar si los efectos fijos de los países y de tiempo pueden o no considerarse iguales se utiliza *el test de máxima verosimilitud* para la redundancia de los efectos fijos, se elige $View\ \rightarrow\ Fixed/Random\ Effects\ Testing \rightarrow\ Redundant\ Fixed\ Effects -$

Likelihood Ratio. Se observan p-valores menores que 0,05 (Figura 35), lo que lleva a afirmar que los efectos fijos de los países y de tiempo son diferentes con un 95% de confianza.

Equation: UNTITLED Workfile: CO...

View Proc Object Print Name Freeze Estimate Forecast Stats Resid

Cross-section Fixed Effects

	CROSSID	Effect
1	1	0.654388
2	2	0.918079
3	3	1.683783
4	4	-3.256250

Figura 33

Equation: UNTITLED Workfile: CO...

View Proc Object Print Name Freeze Estimate Forecast Stats Resid

Period Fixed Effects

	DATEID	Effect
1	1/1/1990	-0.492487
2	1/1/1991	-0.401689
3	1/1/1992	-0.304957
4	1/1/1993	-0.326769
5	1/1/1994	-0.229955
6	1/1/1995	-0.184755
7		

Figura 34

Equation: UNTITLED Workfile: COBBDOUGLAS::...

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Redundant Fixed Effects Tests
Equation: Untitled
Test cross-section and period fixed effects

Effects Test	Statistic	d.f.	Prob.
Cross-section F	17.892324	(3,54)	0.0000
Cross-section Chi-square	55.212136	3	0.0000
Period F	7.884113	(19,54)	0.0000
Period Chi-square	106.251678	19	0.0000
Cross-Section/Period F	22.970444	(22,54)	0.0000
Cross-Section/Period Chi-square	187.023273	22	0.0000

Cross-section fixed effects test equation:
Dependent Variable: LOG(AGRIPRO)
Method: Panel Least Squares
Date: 11/07/14 Time: 13:44
Sample: 1990 2009
Periods included: 20
Cross-sections included: 4
Total panel (balanced) observations: 80

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	19.19683	0.346866	55.34371	0.0000
LOG(AGRICAP)	-1.872818	0.034585	-54.15114	0.0000
LOG(AGRITIERRA)	0.533856	0.012808	41.68272	0.0000
LOG(AGRITRAB)	0.751004	0.011607	64.70167	0.0000

Effects Specification

Figura 35

Según los resultados de las Figuras 29 y 30, la ecuación del modelo de efectos fijos ajustado es la siguiente:

$$Log(AGRIPRO)_{it} = 15{,}06 - 0{,}92 * \mathrm{Log}(AGRICAP)_{it} + 0{,}33 * Log(AGRITRAB)_{it} + 0{,}46 * Log(AGRITIERRA)_{it} + 0{,}65 * d_1 + 0{,}92 * d_2 + 1{,}69 * d_3 - 3{,}26 d_4 - 0{,}49 * F_{90} - 0{,}40 * F_{91} + \cdots + F_{09} * 0{,}37$$

($d_i = 1$ para observaciones del país i y vale $d_i = 0$ en caso contrario, de la misma forma $F_t = 1$ para observaciones del año t y $F_t = 0$ en caso contrario)

Asimismo este modelo presenta normalidad y homocedasticidad. Para comprobar la normalidad se utiliza el test de Jarque Bera. Mediante $View \rightarrow Residual\ Tests \rightarrow Histogram - Normality\ Test$ se observa que el p-valor del contraste de normalidad de

Jarque Bera es superior a 0,05; lo que indica residuos normales con una confianza del 95%. Para comprobar si los residuos son homocedásticos se procede como en la Figura 21 y 22, comprobándose que los residuos presentan varianza constante y son homocedásticos.

El modelo general de efectos aleatorios presenta los residuos $U_{it} = u_i + v_t + w_{it}$ siendo todas las componentes variables aleatorias (en efectos fijos u_i, v_t eran deterministas). Pero habitualmente en efectos aleatorios suele prescindirse de los efectos temporales, teniendo sólo en cuenta u_i a los que se denomina efectos individuales.

A continuación estime el panel con efectos aleatorios y compruebe que el mejor ajuste de los datos es el panel de efectos fijos de corte transversal y de tiempo. Además el panel de efectos aleatorios no es adecuado ya que el test de Hausman, $View \rightarrow Fixed/Random\ Effects\ Testing - f\ Correlated\ Ramdom\ Effects - Hausman\ Test$, presenta un p-valor menor que 0,05 lo que lleva a afirmar que la hipótesis de que los efectos individuales están incorrelacionados con las variables explicativas debe de ser rechazada. Por lo tanto el modelo de efectos aleatorios no es adecuado.

Se llega a la conclusión de que el mejor ajuste de panel es el que considera efectos fijos de sección cruzada y de tiempo.

Bibliografía

- Mayorga, M. y Muñoz S., E. (2000) *La técnica de datos de panel, una técnica para su uso e interpretación*. Departamento de Investigaciones Económicas. BCCR. [Recuperado de www.bccr.fi.cr/investigacioneseconomicas/metodoscuantitativos/Tecnica_datos_panel,_una_guia_para_su_uso_e_interpretacion.pdf, noviembre 2014)
- Perez Lopez, C. (2006). *Problemas Resueltos de Econometría*. Editorial Thomson Paraninfo.