



Universidad de San Andrés
Departamento de Economía
Licenciatura en Economía

***Testeando simetría y kurtosis sobre los residuos de
un modelo de datos en panel agrícola.***

Autor: Juan Martín Borgnino

Legajo: 23281

Mentor: Walter Sosa Escudero

Buenos Aires, Argentina 2014

Testeando simetría y kurtosis sobre un modelo de datos en panel agrícola.¹

Juan Martin Borgnino
Universidad de San Andrés
Noviembre 2014

Síntesis

En el siguiente trabajo implementamos un test de normalidad sobre los residuos de un modelo de datos en panel agrícola. Iniciamos con una descripción de las razones teóricas y metodológicas que motivan a testear la distribución de los residuos. Continuamos con una explicación del funcionamiento y mecánica del test a implementar. Luego corremos nuestro modelo, computamos los errores del mismo y testeamos normalidad. De nuestro test obtuvimos resultados interesantes. Por un lado, rechazamos la hipótesis nula de distribución normal para el componente de error a nivel departamento. De esto inferimos que posiblemente la distribución de la calidad de la tierra no sea normal. Por el otro lado, no podemos rechazar la hipótesis nula de distribución normal en el término remanente. Esto podría estar indicando lo siguiente: las proyecciones climáticas de temperaturas y precipitaciones son generadas con distribuciones normales; la distribución de las expectativas de inflación de los productores parecen distribuirse normalmente. Esto significa que juzgan poco probable que se llegue a niveles de hiperinflación o deflación. Por último, también creemos que las expectativas sobre la tasa de retención se distribuye normalmente. En otras palabras, son pocos los productores que consideran probable que la tasa de retención aumente o se reduzca excesivamente.

¹ Le agradezco a mi mentor, Walter Sosa Escudero, por ayudarme en un momento difícil para el. Le agradezco a mis hermanos por su cariño y alegría. Les agradezco a mi papa y a mi mama por su amor y apoyo incondicional, por su ejemplo y por todo su sacrificio por mi. Con formar una familia con la mitad del amor que ellos formaron, ya me considero exitoso.

1. Introducción

La distribución Normal o Gaussiana es una de las distribuciones de probabilidad continuas más conocidas y de mayor relevancia. Debe en buena medida su gran utilidad al teorema central del límite el cual establece que, en condiciones muy generales, la suma de n variables aleatorias independientes tiene una distribución que se aproxima a la distribución gaussiana. Esto hace que juegue un rol fundamental en procesos de inferencia estadística, dado que numerosas distribuciones muestrales tienden a la normal a medida que crece la cantidad de datos. Asimismo, numerosos métodos y resultados matemáticos dependen del supuesto de normalidad.

Como bien dijimos, muchas variables tienden a la distribución normal conforme crece el tamaño de la muestra. Por lo tanto, dicha distribución proporciona una adecuada modelización del comportamiento de una amplia cantidad de variables. Ejemplos clásicos de variables distribuidas normalmente son: los resultados en exámenes de clases grandes, datos meteorológicos, las alturas de las personas, entre otras. Sin embargo, esto no significa que toda variable pueda ser representada por esta distribución. En muchos casos, la distribución gaussiana puede proveer una buena aproximación en torno a la media, pero no ser precisa ó fidedigna para valores en los extremos. Por ejemplo, Aparicio & Estrada (1997) llaman la atención acerca de que los retornos de acciones de mercados europeos se ajustan mejor a una distribución con colas gordas. En otras palabras, encuentran que eventos extremos son más probables de suceder de lo que predeciría la campana de Bell. Para una empresa de seguros, esto no es un dato menor, ya que dependiendo de la probabilidad de suceso de pérdidas extremas, será el valor que asigne a sus primas.

2. Objetivo del Trabajo

El objetivo del siguiente trabajo es implementar un novedoso test de hipótesis de normalidad de los residuos de un modelo de panel a datos pertenecientes al sector agrícola de la provincia de Santa Fe, Argentina. Es decir, son dos puntos claves los que motivan la siguiente investigación.

En primer lugar, la posibilidad de realizar una investigación utilizando uno de los primeros tests de normalidad de panel. Es innovador porque solo recientemente se lograron resolver las cuestiones de tests de normalidad en panel, mientras que para series de tiempo y datos de corte transversal existe una amplia literatura ya desarrollada. Como se verá, resulta importante testear la normalidad de los residuos por al menos 3 razones básicas. Primero, para muestras finitas es necesario que se cumpla el supuesto de normalidad para poder realizar inferencia estadística mediante t-tests de los estimadores. Segundo, precisamos del supuesto de normalidad para garantizar que los estimadores de máxima verosimilitud sean buenos. Por último, contar con errores normales implica la existencia de pocos outliers, es decir una kurtosis más cercana a 3, y simetría en los errores. Consideramos que resulta una contribución interesante al campo de las investigaciones empíricas de economía y esperamos que este trabajo motive a otros a seguir el mismo procedimiento.

En segundo lugar, hemos decidido aplicar el test a los residuos de un modelo de panel agrícola que se utiliza para estudiar los cambios en la oferta de los principales cultivos por variables como precios, tecnologías agropecuarias, condiciones climáticas, entre otras. Si bien existe una amplia literatura desarrollada sobre la forma en que se debería especificar la distribución de la producción agrícola para Estados Unidos, no existen estudios de este tipo para Argentina. Asimismo los estudios enfocados en Estados Unidos obtienen resultados opuestos y poco claros entre sí. A esto se le debe sumar el hecho de que la producción agrícola es uno de los factores más importantes que afecta a nuestra economía. Por estas razones, consideramos relevante aplicar el test a datos agrícolas. En particular se utilizará el modelo de Nerlove para desarrollar un modelo de panel, se computaran los errores y sobre estos se testeará la normalidad.

El trabajo se divide en dos partes principales. La primera trata fundamentalmente sobre el aspecto teórico de la presente investigación. En particular, tratamos el rol que juega la normalidad en econometría, donde se explica la necesidad

del supuesto de normalidad para realizar inferencia estadística; como la normalidad de errores especifica la distribución condicional de la variable explicada; el efecto de la normalidad sobre los estimadores y algunos detalles sobre el teorema central del limite. La investigación continua con una descripción sobre los tests de normalidad en general en econometría y una explicación con más detalle sobre la mecánica del test que se implementará. Finalmente, presentaremos el modelo que correremos, denominado modelo de Nerlove. La segunda parte se enfoca en el aspecto empírico. Aquí presentamos los datos agrícolas que utilizaremos, junto con una descripción de las variables del modelo. Seguidamente presentamos los resultados de nuestra regresión. Finalmente, se encuentra la parte más importante del presente trabajo donde implementamos el test y analizamos los resultados. En último lugar, realizamos una conclusión de la investigación con una síntesis sobre lo realizado y una breve recapitulación de los resultados.

3. Normalidad en Econometría²

En econometría, la distribución normal juega un rol importante en los residuos de los modelos por al menos 3 razones básicas. En primer lugar, el supuesto es vital en el contexto de los modelos de regresión y muestras finitas para poder realizar inferencia estadística. Segundo, la normalidad de los errores permite conocer con exactitud la distribución condicional de la variable explicada. Tercero, la normalidad es un supuesto que garantiza ciertas propiedades de eficiencia de los estimadores, tales como el buen funcionamiento de los estimadores de máxima verosimilitud. Por último, la normalidad provee errores con pocos outliers y errores simétricos. A continuación desarrollaremos con más detalle cada punto.

Primero, si bien los supuestos clásicos de $E(\mu) = 0$ y $V(\mu) = \sigma^2$ son importantes ya que permiten describir la precisión de los estimadores de MCO, entre otras cosas, no son suficientes para realizar inferencia estadística. Para poder trabajar con tests de hipótesis con estadísticos T y F, necesitamos conocer la distribución

²La gran mayoría de los argumentos, ideas y explicaciones presentes en la sección 3 están basados en Wooldridge (2009).

muestral de los estimadores. Es importante destacar que, sin incluir el supuesto de normalidad, bajo los supuestos clásicos los estimadores de MCO pueden tomar cualquier distribución mientras cumplan con la condición de Esperanza cero y varianza constante.

La distribución muestral de los estimadores condicionada en el valor de las variables independientes, depende exclusivamente de la distribución de los residuos. Esto es, conociendo la distribución de los residuos, podemos conocer la distribución condicional de los estimadores y así realizar distintos ejercicios de inferencia estadística. Por ende, para definir de forma exacta la distribución de los estimadores, se supone que los residuos están distribuidos normalmente. Así se llega a un primer resultado importante:

$$1) \hat{\beta}_j \sim \text{Normal} [\beta_j, \text{Var}(\hat{\beta}_j)]$$

$$2) (\hat{\beta}_j - \beta_j) / ds(\hat{\beta}_j) \sim \text{Normal}(0,1)$$

Ahora bien, conociendo con exactitud la distribución de los estimadores se puede obtener el siguiente resultado:

$$3) (\hat{\beta}_j - \beta_j) / se(\hat{\beta}_j) \sim t_{n-k-1}$$

El mismo se obtiene a partir del resultado en la expresión 2), solo que utilizamos la distribución t y no la normal porque reemplazamos σ presente en $ds(\hat{\beta}_j)$ por $\hat{\sigma}$. Este resultado es importante porque permite realizar tests de hipótesis que involucren a los estimadores.

Cuando se trabaja con este tipo de tests, es importante tener bien presente que se esta queriendo estudiar. Por ejemplo, si quisiéramos testear que un determinado coeficiente es igual a cero, estaríamos preguntándonos en realidad si, luego de controlar por las n-1 variables explicativas, la variable enésima no tiene efecto alguno sobre la variable explicada.

Otra ventaja de los supuestos clásicos es que permite la construcción de intervalos de confianza que establecen rangos de posibles valores que los parámetros de la población pueden tomar bajo una determinada probabilidad. Utilizando el resultado dado en 3), podemos construir un intervalo de confianza para un estimador dado, que muestre el rango de valores que este puede tomar con una determinada probabilidad.

Una segunda consecuencia importante del supuesto gaussiano es que también define la distribución condicionada de y . Son 3 condiciones las que se requieren para justificarlo. Primero, y es una combinación lineal de sus variables explicativas, de sus estimadores y del término de error. Segundo, los estimadores de MCO ya vimos que son normales si se cumple el supuesto de normalidad en los errores. Tercero, la suma de variables normales produce una variable con distribución normal. En consecuencia, si se condiciona en las variables explicativas, y no es más que la suma lineal de variables normales y , por ende, su distribución es gaussiana. Puede sintetizarse de la siguiente manera:

$$y/x \sim \text{Normal}(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k, \sigma^2)$$

En tercer lugar, como se explicó anteriormente, la normalidad de los residuos también tiene efectos sobre la eficacia de los estimadores. Son dos puntos claves los que se quieren destacar aquí. Primero, bajo el resto de los supuestos clásicos, el teorema de Gauss-Markov prueba que los estimadores de MCO son los estimadores lineales insesgados con menor varianza. Sin embargo, el supuesto de distribución gaussiana otorga a los estimadores una propiedad de eficacia más fuerte aún: los convierte en los estimadores insesgados con menor varianza entre todos los tipos de estimadores, no solo los lineales. Esto es así, porque cuando los errores están distribuidos normalmente, el estimador de MCO es equivalente al estimador de máxima verosimilitud. En general, este estimador está bien definido bajo el supuesto de normalidad, pero puede requerir métodos de estimación más robustos o no ser útil en caso de que la distribución de los errores se desvíe de la normal. Mientras que MCO no requiere de normalidad para generar un estimador insesgado, el procedimiento de máxima verosimilitud si lo requiere.

Por último, conocer la distribución de los errores puede ser una pregunta relevante per se. Al econometrista ó investigador le puede interesar conocer la

distribución de los errores para saber si existen outliers que están haciendo crecer la varianza de forma innecesaria. Asimismo, le permite saber si los errores son simétricos o si están sesgados hacia algún lado, proveyéndole información sobre la ubicación (derecha o izquierda respecto a la media) de los valores más extremos. Se trata de información relevante per se y que provee detalles sobre la distribución de las variables explicativas que no se han incluido en el modelo.

El argumento que justifica el supuesto de la distribución normal de los residuos hace uso del Teorema Central del Limite. Este teorema establece que, en condiciones muy generales, la suma de n variables aleatorias independientes tiene una distribución que se aproxima a la distribución gaussiana. Como los residuos, en general, se componen de una suma de variables independientes que afectan a la variable explicada, se puede invocar a dicho teorema para concluir que los errores tienen distribución normal. Si bien el argumento es robusto, depende casi exclusivamente de que los residuos se compongan de variables que afectan a y de forma separada y aditiva. Sin embargo, si los errores son una función cuadrática, por ejemplo, de las variables no incluidas en el modelo, el teorema no podría aplicarse.

Según los autores Zeckhauser & Thompson (1970), hay varias condiciones que pueden volver inoperativo al teorema central del límite y, en efecto, violar a la hipótesis de normalidad de los errores. El número de errores produciendo contribuciones significativas al término de error es usualmente bajo y puede que estos factores tratados como variables aleatorias sean dependientes. A menos que estos factores estén distribuidos normalmente, no se agregaran para producir una distribución normal en el término del error. Agregan que el teorema de Gauss-Markov, implícitamente afirma que la varianza es una medida o métrica valiosa para medir la eficiencia o ajuste de una regresión lineal. Los autores sostienen que el valor y la significatividad de esta medida va disminuyendo a medida que la distribución de los errores diverge de la distribución normal. Prueban esto y concluyen entonces que para muchos casos de interés, cualquier medida de varianza se vuelve inútil o incorrecta como medida de la eficiencia de una regresión. Por ende, conocer la distribución de los residuos nos provee información sobre la confiabilidad de la varianza como medida de eficiencia de los estimadores.

Como podemos ver, existen múltiples razones que prueban la necesidad de testear la normalidad de los errores. Algunas razones son más fundamentales y corresponden a necesidades metodológicas que requieren del supuesto de normalidad

de forma obligatoria. Además, si bien el teorema central del límite es robusto para justificar el supuesto de normalidad, existen condiciones que pueden volver inoperativo a dicho teorema. En síntesis, testear la normalidad de los errores es importante y debería ser un procedimiento estándar en cualquier modelo de regresión.

3.1 Tests de normalidad

Como se explicó en la introducción, el supuesto de normalidad es uno de los más comunes para muchos procedimientos estadísticos. En consecuencia, existe hoy en día una amplia literatura desarrollada sobre el tema. Se han diseñado un gran número de tests para evaluar si muestras de datos y variables se distribuyen de acuerdo a la campana de Gauss. Asimismo, existen procedimientos gráficos, tests de outliers y de bondad de ajuste que también pueden ser empleados para corroborar la normalidad de una muestra de interés. Los teóricos del tema también se han ocupado de estudiar cual es el efecto de la falta de normalidad en los procedimientos que suponen esta distribución para su desarrollo.

Probablemente el test de normalidad más utilizado y uno de los pioneros en el tema es el Jarque Bera. Se trata de un test de bondad de ajuste que prueba si los coeficientes de kurtosis y simetría de una muestra de datos son iguales a 3 y 0, respectivamente. Es decir, compara estos coeficientes con los que se obtendría en una muestra de datos distribuidos normalmente. Si ambos coeficientes de la muestra no se desvían significativamente de estos valores, entonces el test concluye que no se puede rechazar la hipótesis de normalidad. El estadístico de Jarque Bera para regresiones es el siguiente:

$$JB = \frac{n - k}{6} \left(S^2 + \frac{1}{4}(K - 3)^2 \right)$$

donde n es el número de observaciones, k el número de regresores y S y K los coeficientes de simetría y kurtosis, respectivamente. Estos dos últimos están definidos de la siguiente manera:

$$S = \frac{\widehat{\mu}_3}{\widehat{\sigma^3}} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^{3/2}}$$

$$K = \frac{\widehat{\mu}_4}{\widehat{\sigma^4}} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\right)^2}$$

A pesar de que existen diversos métodos y tests para poner a prueba la normalidad, en particular para datos de serie de tiempo y de corte transversal, la literatura es bastante más acotada para datos de panel. Los autores Galvao, Montes-Rojas, Sosa-Escudero & Wang (2013) justifican dicha escasez argumentando que es más complejo testear la normalidad de los errores en datos de panel. Esto último se debe a que los errores en estos modelos están formados por dos componentes diferenciados. Para los modelos de datos de panel de one-way component error model, los residuos se modelan de la siguiente manera:

$$u_{it} = \alpha_i + \varepsilon_{it}$$

donde α_i representa las variables individuales no observadas que afectan a y , pero que no varían en el tiempo; ε_{it} representa todas las variables no incluidas en el modelo que afectan a y y que varían en tiempo e individuo. En efecto, al problema de detección de desvío de normalidad se le suma el interrogante sobre cual de los dos términos es el responsable por dicho desvío.

A pesar de dichas complicaciones, testear normalidad en datos de panel, al igual que para datos de corte transversal, resulta de gran relevancia. Los autores Galvao, Montes-Rojas, Sosa-Escudero & Wang (2013) han desarrollado un test que permite poner a prueba la simetría y kurtosis de los errores pertenecientes a estos modelos. Los tests hacen uso de las transformaciones within y between de los residuos de MCO. Prueban que bajo la hipótesis nula, luego de estandarizar y centralizar, los estadísticos tienen distribución asintóticamente normal. Los estadísticos para probar dichas propiedades son los siguientes:

$$\frac{\sqrt{N}(\widehat{SK}_v - s_v)}{\sqrt{\Omega_v^s}} \xrightarrow{d} N(0,1), \quad \text{donde } \Omega_v^s = NVar(\widehat{SK}_v);$$

$$\frac{\sqrt{N}(\widehat{SK}_u - s_u)}{\sqrt{\Omega_u^s}} \xrightarrow{d} N(0,1), \quad \text{donde } \Omega_u^s = NVar(\widehat{SK}_u);$$

$$\frac{\sqrt{N}(\widehat{KU}_v - k_v)}{\sqrt{\Omega_v^k}} \xrightarrow{d} N(0,1), \quad \text{donde } \Omega_v^k = NVar(\widehat{KU}_v);$$

$$\frac{\sqrt{N}(\widehat{KU}_u - k_u)}{\sqrt{\Omega_u^k}} \xrightarrow{d} N(0,1), \quad \text{donde } \Omega_u^k = NVar(\widehat{KU}_u);$$

Por ende, resulta simple implementar este test para kurtosis y simetría ya que se puede usar un t test para realizarlo. No se podrá rechazar la hipótesis de normalidad en los errores únicamente cuando no se pueda rechazar la hipótesis nula para estos cuatro estadísticos. Igualmente, los autores explican que si bien la matriz de varianza-covarianza de los estadísticos son acotadas por supuestos iniciales, su derivación es compleja. Por lo tanto, para implementar el test se utiliza un procedimiento de bootstrap. Esta metodología es utilizada justamente para medir ciertas propiedades de estimadores mediante la generación de numerosas muestras de datos diferentes que provengan de una distribución aproximada a la que se tiene. Es decir, consiste en reproducir un gran número de muestras con una distribución lo más aproximada posible a la original y en cada una de ellas calcular la propiedad que nos interesa computar de nuestro estimador.

Además de contar con la ventaja de ser simple de implementar, el test permite al investigador conocer dos cuestiones. Primero, detectar la causa del desvío de normalidad, ya sea por asimetría ó kurtosis. Una muestra con distribución normal tiene asimetría igual a cero y kurtosis igual a 3. La segunda cuestión importante es que el test permite localizar al causante de la falta de normalidad. Es decir, permite poner a prueba por separado ó conjuntamente la simetría y kurtosis del componente de error individual y del remanente. El test que se rechace indicará la causa y el componente responsable de la falta de normalidad.

3.2 Presentación del caso

Como se describió en la introducción, consideramos relevante implementar el test de normalidad de datos en panel a un modelo agrícola. Los economistas agropecuarios llevan tiempo buscando la forma de especificar dicha distribución. Lograr una adecuada modelización de la producción del sector agropecuario es importante por varias razones. En primer lugar, Harri, Erdem, Coble & Knight (2009) argumentan que afecta las decisiones de maximización de beneficio de los productores. Según las expectativas que tengan estos sobre su producción, decidirán su inversión en los recursos productivos como semillas, maquinarias, mano de obra y fertilizantes, entre otros. Asimismo, también tiene repercusiones sobre la elección de que proporciones de la tierra utilizarán para cada tipo de cultivo. En segundo lugar, tiene consecuencias importantes sobre las empresas aseguradoras, dado que estas utilizan la distribución estadística para modelar su exposición al riesgo y así definir la prima que cobrarán sobre los seguros que provean a los productores. En tercer lugar, la distribución de la producción agropecuaria también es importante para el establecimiento de planes gubernamentales que protejan a los productores. Xu, Hart, Smith & Meyers (1995) explican que los gobiernos cuentan con programas de seguros y paquetes de compensación para proteger los ingresos de los productores en caso de pérdidas inesperadas. Este tipo de programas hace uso de información estadística para definir montos y porcentajes.

Existen varios riesgos en torno a la producción agropecuaria producto de las variables económicas y biofísicas del contexto en que operan los productores. Estos riesgos pueden tener serios efectos sobre las decisiones de inversión futura y el desarrollo de los productos agrícolas.

En particular, la incertidumbre en el retorno de los productores esta dada en buena medida por la brecha de tiempo que existe entre las decisiones de producción y el momento en que el productor recibe la compensación por su producto. Es en este intervalo de tiempo, en el que distintos factores determinan el valor de la cosecha. Entre estos factores se encuentra el clima, el estado de salud de los cultivos, cambios en los mercados agropecuarios (preferencias, por ejemplo) y un numero de factores macroeconómicos (trabas a las importaciones, subsidios, impuestos, entre otros). Variaciones en estos factores es lo que genera riesgo en los 3 determinantes del

retorno de los productores agropecuarios: volumen de producción, costo productivos y el precio de sus cultivos.³

Por último, existen ciertas variables que sabemos que estarán incluidas en nuestro término de error para las cuales resulta interesante conocer si su distribución se desvía de la normalidad. Dado que, como veremos, nuestro modelo trata sobre la decisión de siembra, existen variables de expectativas inflacionarias y políticas que afectan nuestra variable explicada. Considerando el contexto económico recesivo, de escasez de divisas y de alza en los precios, contar con información adicional sobre las expectativas de los agentes resulta interesante.

En definitiva, conocer la distribución de la producción agropecuaria no es un detalle menor. Su forma afecta importantes decisiones económicas que competen desde productores, hasta empresas aseguradoras y programas estatales. Por estas razones, un gran número de economistas y estadistas aplicados se han dedicado a producir una vasta literatura sobre el tema. Sin embargo, han fracasado en llegar a un consenso sobre cómo especificar la distribución de la producción agropecuaria para Estados Unidos⁴. Algunos estudios han rechazado la hipótesis de normalidad, otros concluyen que cuenta con asimetría positiva y algunos que la asimetría es negativa. En buena medida, la divergencia de resultados se puede deber a que muchos de estos estudios utilizan diferentes datos y tests de normalidad. Asimismo, dichas investigaciones usan datos de un número acotado de zonas geográficas, por lo que sus hallazgos tienden a ser poco robustos. En síntesis, la literatura nos deja con resultados inconclusos y preguntas sin respuesta.

³ Los detalles de dicho párrafo pueden verse mejor en un reporte producido por el gobierno escocés sobre riesgos y estrategias agrícolas en el 2010.

⁴ Véase Harri, Erdem, Coble & Knight (2009) para contar con un reseña sobre los diferentes trabajos que trataron dicha cuestión.

3.3 El modelo

El modelo a estimar es conocido como “Modelo Nerloviano”, nombre derivado de su autor Marce Nerlove. Es un modelo conocido en la literatura de Economía Agrícola que se emplea principalmente para estudiar la respuesta de la oferta de ciertos cultivos ante cambios en precios de cultivos sustitutos, costos de producción y otras variables exógenas. Como se explicó anteriormente, los productores deben decidir su nivel de producción en base a los precios que estos esperan recibir en el futuro una vez que su cosecha este lista. Por ende, el modelo nerloviano se construye para examinar la reacción de los cambios de nivel de producción ante los precios esperados. Puede utilizarse para explicar el cambio en variables de oferta como hectáreas sembradas ó volumen de producción. El modelo consiste de 3 ecuaciones principales:

$$4) A_t^d = \alpha_1 + \alpha_2 P_t^e + \alpha_3 Z_t + u_t$$

$$5) A_t - A_{t-1} = \delta(A_t^d - A_{t-1}) + v_t$$

$$6) P_t^e - P_{t-1}^e = \gamma(P_{t-1} - P_{t-1}^e) + w_t$$

donde,

A_t^d = volumen de cultivo deseado a producir en el tiempo t.

P_t^e = precio esperado en el tiempo t del cultivo y de otros cultivos competidores.

A_t = volumen de cultivo realmente producido en t.

Z_t = otros factores exógenos que afectan el volumen producido en t.

Las ecuaciones 4), 5) y 6) sirven para modelar los factores que determinan los volúmenes de producción y la formación de expectativas del productor. En primer lugar, la ecuación 4) establece que el volumen de cosecha que el productor desea producir es una función de los precios relativos esperados y un número de variables exógenas, como el clima, fertilizantes, etc. La ecuación 5), sirve para explicitar que no es posible un ajuste total al nivel deseado de volumen de producción en el corto plazo.

El ajuste real de producción será solamente una fracción del volumen deseado. Por último, la ecuación 6) es una expresión que especifica como los agentes forman sus expectativas de precios basándose en los precios actuales y pasados, dado que el precio que los productores esperan en el tiempo t es desconocido. Los productores ajustan sus expectativas mediante una fracción de la diferencia entre el precio actual en $t-1$ y sus expectativas de precios de $t-1$. Es decir, utilizan su experiencia pasada para formar expectativas.

Como las variables A_t^d y P_t^e son desconocidas, utilizamos las 3 ecuaciones para llegar a una única ecuación que no dependa de estas dos variables.

$$7) A_t = \pi_1 + \pi_2 P_{t-1} + \pi_3 A_{t-1} + \pi_4 A_{t-2} + \pi_5 Z_t + e_t$$

La ecuación 7) es el modelo que podemos estimar especificado por las ecuaciones 4), 5) y 6) ya que todas las variables son conocidas. Se trata de un modelo que depende del rezago de 2 período anteriores de la variable dependiente, del precio del cultivo y de cultivos competidores del periodo anterior y de otras variables exógenas que luego detallaremos.

Como bien explican Yu, Liu & You (2010), la decisión de producción de los agricultores depende de factores monetarios y no monetarios. Por un lado, entre los factores monetarios, se destaca el precio futuro en los mercados a términos y los costos de producción. En el momento de siembra, los productores analizan cuál será su inversión y en que tipo de cultivo considerando los precios en la Bolsa de Comercio de Rosario. Por último, consideran los costos de producción. Normalmente, acuerdan pagar los costos de producción con las cosechas futuras, por ende se congela el precio de ese mes. La teoría sugiere que al subir el precio de los cultivos competidores aumentará el incentivo a cambiar la producción del cultivo actual por la del cultivo sustituto. Por el otro lado, los factores no monetarios incluyen factores tecnológicos exógenos que aumentan la productividad y factores físicos como lluvias, riegos, inversión y educación.

4. Datos y agricultura en Argentina

Para este estudio, nos concentraremos en el área sembrada de soja para 3 departamentos de la provincia de Santa Fe, Argentina durante los años 1995 y 2008. Es importante para esto entender el ciclo de producción de la soja. En nuestro país, los meses de siembra óptimos para este cultivo son Octubre y Noviembre. Existe igualmente la llamada “soja de segunda fecha de siembra” la cual refiere a la soja que se siembra en Diciembre. Sin embargo, en nuestro país la mayor superficie de esta oleaginosa se siembra durante los meses previos, dado que estas siembras suelen obtener los máximos rindes. La cosecha normalmente ocurre durante los meses de Marzo y Mayo, siendo este último el mes que mayor actividad suele presentar.

Para realizar nuestro análisis, la unidad geográfica de estudio es el área en el entorno de cada estación meteorológica de dicha provincia. En particular, nos concentraremos en los departamentos de General Obligado, Castellano e Iriondo. Elegimos estos tres departamentos porque obtuvimos datos de variables meteorológicas de las estaciones en la ciudad de Iriondo, Rafaela y Oliveros.

En este estudio, utilizamos 7 variables exógenas para explicar el área sembrado de la soja. Si bien vimos que el modelo de Nerlove utiliza el precio de las cosechas pasadas para estimar el precio esperado de los agricultores, consideramos que un mejor estimador es el precio, en el momento de siembra, del futuro con vencimiento en momento de cosecha. Los precios futuros son, según la teoría económica, los mejores estimadores insesgados de los precios spot futuros. Esta modificación del modelo original de Nerlove no solo esta basada en la teoría económica sino también en la recomendación de economistas agrarios e ingenieros agrónomos⁵. Obtuvimos la serie histórica de precios futuros del mercado a término de Buenos Aires, llamado Matba. A estos precios se les aplicó dos tratamientos. Primero, dado que las estaciones meteorológicas están a diferentes distancias del puerto de Rosario, computamos el costo de flete promedio que enfrentan los productores ubicados en cada estación. Para esto, utilizamos datos provenientes del sitio de FEDEEAC (Federación Argentina de Entidades Empresarias del Autotransporte de Cargas) que registra el costo por tonelada y kilometro del transporte de cereales y oleaginosas. Este precio no es en absoluto el valor neto que recibe el productor por

⁵ El ingeniero agrónomo Gustavo Richiger y el economista agrario Victor Breschia consideraron que era un cambio favorable y realista al modelo.

tonelada, pero es una aproximación a lo que cada productor recibe en cada estación. Finalmente, deflactamos estos precios por el Índice de Precios al Por Mayor con año base en 1995=100. Además del precio de la soja, también incluimos el precio del maíz y del girasol por ser el sustituto más cercano. A estos dos precios le aplicamos el mismo calculo que al precio de la soja. La teoría económica sugiere que en caso de aumentar el precio de un bien competidor, la producción de la soja debería verse afectada negativamente.

Asimismo, utilizamos dos variables climáticas. Para cada estación obtuvimos datos mensuales de precipitaciones y temperaturas. Asumimos que los productores deciden cuantas hectáreas sembrar en base a la humedad del suelo, la cual depende de las precipitaciones y, en menor medida, de las temperaturas de los meses previos. Estos datos climáticos fueron obtenidos del INTA (Instituto Nacional de Tecnología Agropecuaria).

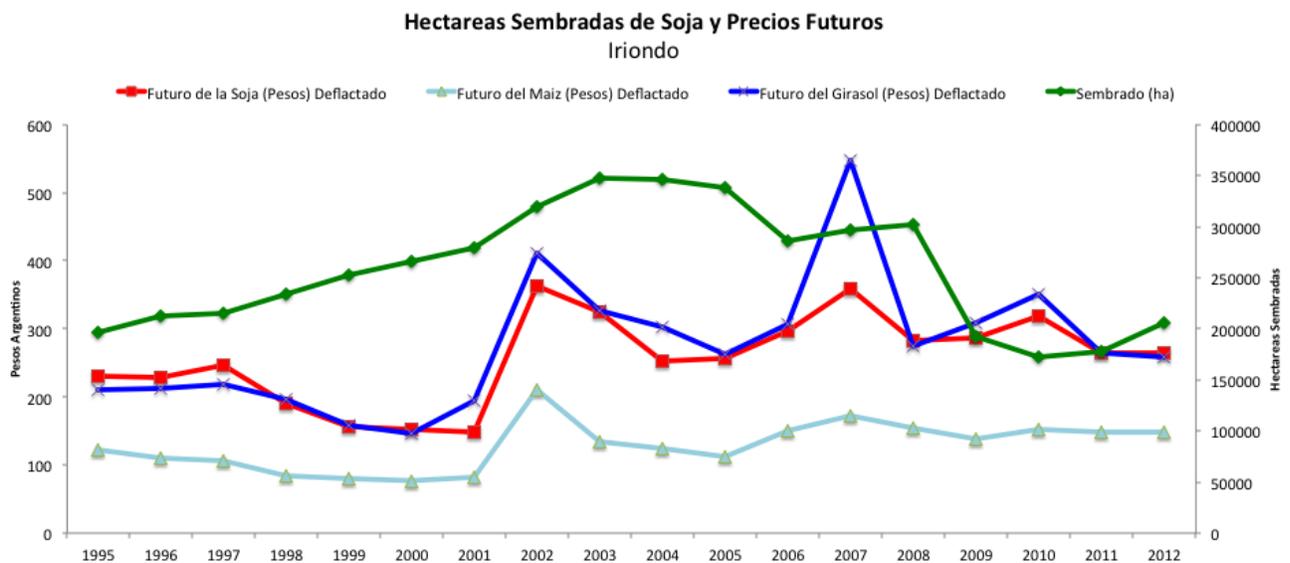
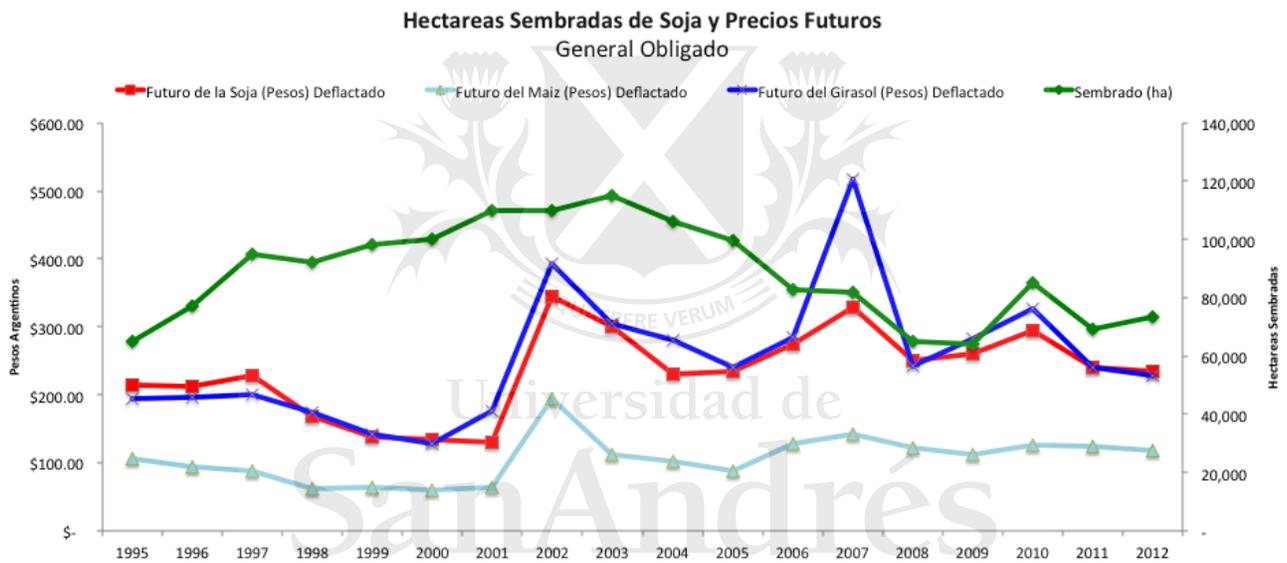
Por último, incluimos 2 variables que captan algunos de los costos de la producción agrícola. Uno de estos, son los costos de labranzas que agrega en una única variable costos de siembra directa, de fertilización y de pulverización terrestre y aérea. Por último, agregamos el costo de un herbicida llamado 2-4 D 100%. Contamos con datos de otros costos, pero decidimos no agregarlos por su baja correlación con las hectáreas sembradas y alta correlación con otras variables del modelo. Para evitar problemas de multicolinealidad, decidimos excluirlas y vimos mejoras significativas de nuestro modelo. Las variables mencionadas de insumos las obtuvimos de la revista Márgenes Agropecuarios que se especializa en la publicación de números del campo. Cabe destacar que todas las variables monetarias fueron deflactadas.

La siguiente tabla resume las variables que hemos explicado y que se utilizaran para realizar nuestras estimaciones:

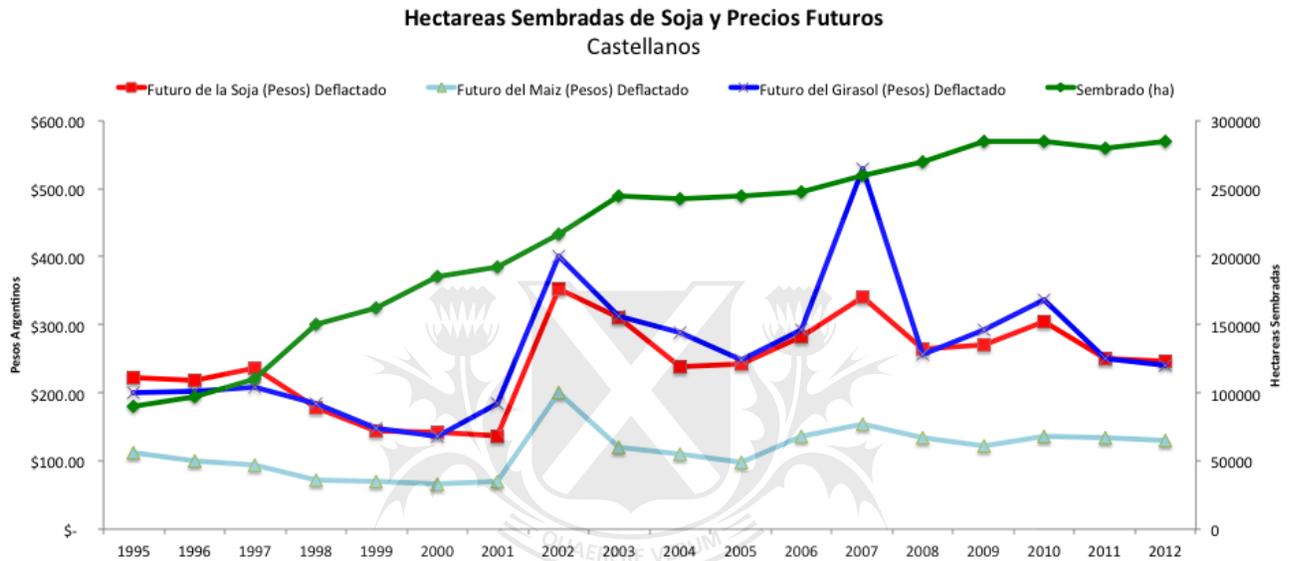
Variables	Descripción
Sembrado	Logaritmo del las hectáreas sembradas de Soja por año en los Departamentos de Gral. Obligado, Castellanos e Iriondo de la Provincia de Santa Fe, Argentina. Fuente: sitio web del SIIA (Sistema Integrado de Información Agropecuaria) perteneciente al Ministerio De Agricultura, Ganadería y Pesca.
Futuro_Soja	Logaritmo del precio promedio del futuro de la soja, en el momento de siembra (20 de Octubre – 10 de Noviembre), con vencimiento en el momento de cosecha (Mayo), descontando costo promedio de flete al puerto de Rosario desde cada Departamento. (Deflactado). Fuente: MATba (Mercado a Término de Buenos Aires).
Futuro_Maiz	Logaritmo del precio promedio del futuro del maíz, en el momento de siembra (Septiembre – Octubre), con vencimiento en el momento de cosecha (Abril), descontando costo promedio de flete al puerto de Rosario desde cada Departamento. (Deflactado). Fuente: MATba (Mercado a Término de Buenos Aires).
Futuro_Girasol	Logaritmo del precio promedio del futuro del girasol, en el momento de siembra (Septiembre – Diciembre), con vencimiento en el momento de cosecha (Marzo), descontando costo promedio de flete al puerto de Rosario desde cada Departamento. (Deflactado). Fuente: MATba (Mercado a Término de Buenos Aires).
Labranzas	Logaritmo de costos de labranzas por año obtenido de coeficientes de la UTA. Incluye fertilizantes, pulverizantes y siembra directa (Deflactado). Fuente: serie histórica de la revista Margen Agropecuario.
2-4 D 100%	Logaritmo de costo de herbicida utilizado para la siembra de la soja (Deflactado). Fuente: serie histórica de la revista Margen Agropecuario.
Temperaturas_Promedio	Temperatura promedio de Agosto, Septiembre y Octubre por región correspondiente. Fuente: INTA.
Precipitaciones_Total	Suma de precipitaciones durante Agosto, Septiembre y Octubre por región correspondiente. Fuente: INTA.

A continuación presentamos un gráfico para cada departamento de las hectáreas sembradas de soja y los precios futuros de la soja, el maíz y el girasol.

Por un lado, para los departamentos de General Obligado e Iriondo, entre 1995 y 1999 a pesar de que los precios fueron en caída, el sembrado creció. Una posible explicación, podría ser que los productores debieron compensar la caída en sus ingresos por los precios, a través de un aumento del volumen de producción. Luego en el 2001, los precios en pesos se disparan por la devaluación. Notar que en el 2002, coincide el pico de sembrado de soja, con el pico de precio del futuro de la soja. Asimismo, en el 2007 se dispara el futuro del girasol y coincide con una caída en el sembrado de la soja. Por último, la caída de la siembra durante el período final puede ser producto de aumentos en los alquileres de los campos (consecuencia de que la rentabilidad de producir soja era muy alta por su precio).



Por el otro lado, en el departamento de Castellanos vemos que el sembrado de la soja creció durante todo el período independientemente de los precios. Una diferencia importante, es que el departamento de Castellanos fue una región de campos históricamente ganaderos. Sin embargo, producto del fuerte aumento de los precios de la soja y una caída importante de los márgenes del sector ganadero, estos campos se volcaron a la agricultura, en detrimento de la ganadería.



Universidad de
San Andrés

4.1 Estadística Descriptiva

En primera instancia se corrió un modelo con más variables que las detalladas en el punto anterior. Se incluyeron 2 variables adicionales para controlar por los costos de producción: costos de sueldos promedio de peones de campo y el de un herbicida llamado glisofato. La siguiente tabla muestra los resultados del modelo con todas las variables.

<i>Variable Explicada: Sembrado</i>	<i>Coefficientes (Desv. Estand.)</i>
<i>Futuro_Soja</i>	0,124 (0,136)
<i>Futuro_Maiz</i>	-0,097 (0,089)
<i>Futuro_Girasol</i>	0,114 (0,056)**
<i>Labranzas</i>	-0,164 (0,208)
<i>2-4 D 100%</i>	-0,155 (0,117)
<i>Glisofato</i>	0,070 (0,031)**
<i>Sueldos</i>	-0,185 (0,044)**
<i>Sembrado_t-1</i>	1,022 (0,102)***
<i>Temperaturas_Promedio</i>	-0,312 (0,419)
<i>Temperaturas_Promedio_2</i>	-0,001 (0,013)
<i>Precipitaciones_Total</i>	-0,001 (0,001)
<i>Precipitaciones_Total_2</i>	2,27e-6 (1,92e-6)
<i>_cons</i>	3,544 (1,996)*
<i>R²: Within</i>	0,933
<i>Between</i>	0,995
<i>Overall</i>	0,980

* p < 0,10; ** p < 0,05; *** p < 0,01

Los resultados no son los esperados. Las variables que más nos interesan, los futuros de la soja y del maíz, no son estadísticamente significativas para explicar el área sembrada y tienen signos contrarios al que esperaríamos. Un R^2 elevado, coeficientes poco significativos y con signos contrarios parecería indicar problemas de multicolinealidad alta. Por lo tanto, nos encontramos con dos posibles soluciones al problema. Primero, extender la serie de datos. Lamentablemente, esto no fue posible porque no se logro conseguir mayor cantidad de datos para todas las series. Por ende, debimos emplear la segunda solución que consistió en realizar un análisis de correlación para detectar que variables explicativas contaban con correlación alta entre si, pero baja correlación con la variable explicada. El objetivo fue intentar de eliminar el problema de multicolinealidad, pero sin por eso tener un problema de omisión de variable relevante.

La siguiente tabla presenta un análisis de correlación para nuestras variables:

Variables	Sembrado	Futuro_Soja	Futuro_Maiz	Futuro_Girasol	Glisofato	Sueldos
Sembrado	1					
Futuro_Soja	0.222	1				
Futuro_Maiz	0.314	0.934	1			
Futuro_Girasol	0.274	0.910	0.877	1		
Glisofato	0.001	0.568	0.600	0.430	1	
Sueldos	0.030	0.106	0.111	0.171	0.109	1

Como se puede ver, el índice de correlación entre el glisofato y el área de sembrado, y los sueldos y el área de sembrado es bajo. Por lo tanto, parecería ser que estas variable no deberían sesgar nuestros estimadores si no las incluimos. Asimismo, estas variables están correlacionadas con los precios futuros. Por ende, vale la pena probar si nuestro modelo mejora al excluir estas dos variables.

<i>Variable Explicada: Sembrado</i>	<i>Coefficientes (Desv. Estand.)</i>
<i>Futuro_Soja</i>	0,366 (0,114)***
<i>Futuro_Maiz</i>	-0,321 (0,040)***
<i>Futuro_Girasol</i>	0,028 (0,039)
<i>Labranzas</i>	-0,429 (0,150)**
<i>2-4 D 100%</i>	0,111 (0,134)
<i>Sembrado_t-1</i>	1,064 (0,097)***
<i>Temperaturas_Promedio</i>	-0,214 (0,539)
<i>Temperaturas_Promedio_2</i>	0,006 (0,017)
<i>Precipitaciones_Total</i>	-0,001 (0,001)
<i>Precipitaciones_Total_2</i>	1,76e-6 (1,16e-6)
<i>_cons</i>	1,973 (3,004)*
<i>R²: Within</i>	0,922
<i>Between</i>	0,994
<i>Overall</i>	0,977

* p < 0,10; ** p < 0,05; *** p < 0,01

Luego de quitar de nuestro modelo las variables de glisofato y sueldos vemos una clara mejora. Primero y principal, nuestro modelo explica en un 97% la variación del área sembrada. Segundo, las variables de mayor interés son estadísticamente significativas y cuentan con el signo que esperaríamos según la teoría económica. Al aumentar el precio del futuro de la soja en el momento de siembra, ceteris paribus, aumenta el área sembrada de este cultivo. Lo contrario sucede con el maíz, dado que se trata del principal sustituto para los productores. No obtuvimos el coeficiente esperado para el girasol, pero tampoco podemos rechazar la hipótesis nula de que el coeficiente de dicha variable sea igual a cero.

Con respecto al coeficiente de las variables de costos, vemos que labranzas es una variable significativa y cuyo coeficiente cuenta con el signo esperado. Dado que las regiones que hemos elegido cuentan con campos de altos niveles de fertilidad, mucho de los costos de fertilizantes y herbicidas que pueden ser importantes en otros sectores, aquí no son pertinentes. La calidad de la tierra hace que no sea necesario para los productores invertir en algunos insumos que sin son requeridos en otros espacios. La variable 2-4D no es significativa.

Asimismo, ninguna de las variables climáticas parece tener efectos sobre el área sembrada. Existen distintas razones que pueden justificar la baja significatividad. Primero, puede ser posible que algunos de los agricultores de estas zonas utilicen sistemas de riego que vuelvan independiente su decisión de sembrado respecto a las lluvias. Segundo, puede suceder que a pesar de que la humedad del suelo sea importante para la siembra, los agricultores sepan que durante los meses de cultivo las lluvias sean suficientes para garantizar buenos rindes. Al tratarse de regiones cuyas condiciones climáticas históricamente han sido propicias para la agricultura, los productores puede ser que no consideren las lluvias previas a la siembra.

Por último, cabe destacar que los resultados sobre el efecto de los precios sobre la decisión de siembra son similares a los que han obtenido Breschia & Lemma (2001). Los autores analizaron el efecto del precio de la soja y del maíz sobre el área sembrada de maíz y obtuvieron 0.36 para el precio del maíz y -0.23 para el precio de la soja. Si bien nuestra variable explicada es la soja y no el maíz, y los autores usaron precios de campañas anteriores en lugar de los futuros, los resultados son cercanos.

En definitiva, esto parece indicar que los resultados de nuestro modelo son coherentes y podemos proceder a implementar el test de normalidad.

4.2 Resultados del Test

Para implementar el test se utilizó un script de Stata que fue empleado por los autores del test a modo de ejemplo para un panel con datos de empresas. Procedimos a tomar el script y a adaptarlo a los datos del presente trabajo. El script se encarga de generar los estadísticos de prueba, de implementar el procedimiento de bootstrap para aproximar las varianzas de los estimadores y luego de computar los p-valores con los cuales decidiremos si rechazamos o no la hipótesis nula en cuestión. Recordemos que el p-valor muestra la probabilidad de obtener el resultado al que se llegó si suponemos que la hipótesis nula es cierta. Por lo tanto, si los p-valores son muy chicos, implica que es poco probable que la hipótesis nula sea correcta. Rechazaremos la hipótesis nula si el p-valor es menor que el nivel de significatividad.

Son 6 tests de hipótesis que se deben realizar si se quiere testear todas las variantes del residuo. Tanto para el componente de error individual como para el remanente se debe testear por separado la simetría y la kurtosis. Asimismo, se debe testear de forma conjunta la simetría y kurtosis para ambos componentes. Esto último es la prueba de normalidad completa sobre el error. Esta flexibilidad para testear en ambos componentes simetría y kurtosis de forma conjunta y de forma independiente es lo que nos permite conocer con exactitud la fuente de desvío de normalidad.

La siguiente tabla presenta los p-valores que obtuvimos para cada test de hipótesis con datos de nuestro modelo agrícola:

Tests de Hipotesis: H_0	P-Valor
Simetría $v_{it} = 0$	0,5698
Kurtosis $v_{it} = 3$	0,0996
Simetría $v_{it} = 0$ & Kurtosis $v_{it} = 3$	0,2193
Simetría $\mu_i = 0$	0,0094
Kurtosis $\mu_i = 3$	0,0272
Simetría $\mu_i = 0$ & Kurtosis $\mu_i = 3$	0,0030

Estos son los resultados más importantes del presente trabajo:

- Con un p-valor de 0.0094, rechazamos la hipótesis nula de simetría en el componente de error a nivel departamento con un nivel de significatividad del 1%.
- Rechazamos la hipótesis nula de kurtosis = 3 en el término de error a nivel departamento con un p-valor de 0.0272 y con un nivel de significatividad del 5%.
- Como es de esperar por los dos puntos anteriores, también rechazamos el test de hipótesis conjunta de kurtosis y simetría para el error individual con un p-valor de 0.0030 y con un nivel de significatividad del 1%. Esto es, rechazamos la hipótesis nula de que el componente de error individual posee distribución normal.
- Con un p-valor de 0.5698, no podemos rechazar la hipótesis nula de simetría = 0 en el componente remanente.
- Para la hipótesis de kurtosis = 3 en el componente remanente obtuvimos un p-valor de 0.0996. Por lo tanto, rechazaremos o aceptaremos la hipótesis nula dependiendo del nivel de significatividad del test. Usualmente se utilizan niveles de significatividad del 1%, 5% y 10%. En caso de optar por los dos primeros, rechazaremos la hipótesis nula. Caso contrario, aceptamos la hipótesis nula.
- Por último, como es de esperar por los dos puntos anteriores, no podemos rechazar la hipótesis nula conjunta de simetría y kurtosis con valores cercanos a los correspondientes a una distribución normal. Esto significa que no podemos rechazar la hipótesis de que el componente de error remanente se distribuya normalmente.

Con estos resultados presentes, podemos analizar que tipos de variables pueden estar presentes en ambos tipos de errores y algunos indicios sobre las mismas. Primero, la asimetría es resultado de la existencia de datos más separados de la media a la derecha o a la izquierda. En otros términos, existen outliers de un solo lado de la media de la distribución y esto produce que una de las colas sea más larga que la otra. Segundo, distribuciones con exceso de kurtosis presentan, a diferencia de asimetría, outliers de ambos lados de la media. Esto genera que valores extremos de ambos lados

de la media sean más probables de suceder de lo que predeciría una distribución normal. Asimismo, el valor de kurtosis puede ser menor a 3. Esto produce que los valores extremos sucedan con menor probabilidad que en una distribución normal.

Por un lado, obtuvimos que el componente de error individual no se distribuye normalmente. Esto es así, tanto por la falta de simetría y por el desvío de kurtosis del valor 3. Dicho término de error esta compuesto por variables que afectan al área sembrada por departamento pero que permanecen constantes en el tiempo. Una posible variable incluida aquí es la calidad de la tierra. Es decir, una variable que capture en que condición están las tierras para el cultivo agrícola. En caso de que la kurtosis fuese mayor a 3, podría estar indicando que para los 3 departamentos existe una mayor proporción de tierras de muy buena y de muy mala calidad; mayor a lo que obtendríamos con una distribución normal. Lo contrario sucedería si el coeficiente de kurtosis fuese menor a 3. Existirían muy pocas tierras de menor y mayor calidad que la media.

Por el otro lado, el término de error remanente para este caso probablemente cuente con una mayor cantidad de variables. Esto es así, porque a diferencia del término de error individual, este error varía en individuo y en tiempo. Recordemos que para dicho error obtuvimos que no podemos rechazar la hipótesis nula de normalidad. Para esto, las variables que lo componen deben ser normales, dado que sino la suma de estas variables no sería normal. Primero, una variable que podría estar incluida aquí son los pronósticos de clima para la región de los productores. Si los productores observan malos pronósticos de lluvia y de temperaturas para los meses de cultivo, pueden decidir reducir el área sembrada. Por lo tanto, parecería ser que las proyecciones de temperaturas y lluvias son generadas a través de distribuciones normales. Sin embargo, si trabajamos con niveles de significatividad del 1% y 5%, rechazaríamos la hipótesis de kurtosis igual a 3. Esto podría indicar que se proyectan precipitaciones extremas, tanto de sequías como de inundaciones. Al ser simétrica la distribución, también indica que las probabilidades de menores y mayores precipitaciones respecto a la media son igual de probables.

Por último, existen dos variables que influyen en la decisión de siembra y están relacionadas con las políticas económicas de argentina de los últimos años. La primera variable son las expectativas inflacionarias. Esta variable afecta la decisión de siembra del productor, debido a que estima y proyecta posibles aumentos en sus costos productivos. Por ejemplo, el salario de los peones y el costo de flete puede sufrir

aumentos y esto impactaría negativamente sobre la rentabilidad del productor. Nuestro resultado indica que la distribución de las expectativas inflacionarias se distribuyen normalmente. Es decir, son pocos los productores que prevén niveles excesivos de inflación o deflación. La segunda variable corresponde a la política de retenciones. En caso de que los productores consideren que las cargas impositivas sobre el valor de su producción se reduzca, aumentaría las hectáreas de siembra. Nuevamente, las expectativas sobre modificaciones a la tasa de retención parecen distribuirse normalmente. En otras palabras, son pocos los productores que consideran que la tasa de retención pueda reducirse ó aumentar a niveles extremos.



Universidad de
SanAndrés

5. Conclusión

Testear la normalidad de los errores en modelos econométricos es importante por varias razones. Primero, el supuesto de normalidad es clave en muestras finitas para poder realizar tests de hipótesis. Segundo, el desvío de normalidad puede afectar la eficiencia de los estimadores, tales como el buen funcionamiento del estimador de máxima verosimilitud. Por último, conocer la distribución de los errores puede ser una cuestión relevante per se y puede brindar información interesante sobre la distribución de las variables que el investigador no ha incluido en su modelo.

Si bien existen numerosos tests de normalidad para datos de serie de tiempo y de corte transversal, escasean tests de este tipo para datos de panel. Una de las razones es su mayor complejidad, producto de que el error esta compuesto por dos términos: uno dependiente del individuo y del tiempo, y otro independiente del tiempo.

El presente trabajo tuvo como objetivo implementar un novedoso test de normalidad de panel. Recientemente Galvao, Montes-Rojas, Sosa-Escudero & Wang (2013) desarrollaron un test que permite testear si la distribución del término de error individual y el remanente se desvían de simetría = 0 y kurtosis = 3. Es decir, permite probar si la distribución es normal. Una ventaja de dicho test es la flexibilidad para testear simetría y kurtosis de forma independiente para cada componente de error por separado. Esto ayuda a detectar, por un lado, la fuente del desvío de normalidad - individual y/o remanente - y, por el otro lado, la causa del desvío de normalidad - asimetría y/o kurtosis menor o mayor a 3. Además, la implementación del test es relativamente sencilla y solo requiere algunos conocimientos básicos de código de Stata.

El test se implemento a un modelo agrícola, llamado modelo de Nerlove, que se utiliza para estimar las elasticidades precio oferta directas y cruzadas de los cultivos de interés. Optamos por implementar el test a dicho modelo porque la distribución condicional del área sembrada de soja puede tener efectos importantes sobre decisiones económicas de productores, de aseguradoras y del gobierno. Asimismo, como vimos existen variables en el término de error que resultan interesantes para la realidad económica argentina.

Los resultados proveyeron información interesante sobre los residuos. Por una parte, rechazamos la hipótesis nula de distribución normal para el componente de error a nivel departamental. De esto inferimos que posiblemente la distribución de la

calidad de la tierra no sea normal. Esto puede significar que es más probable encontrar tierras de alta y baja calidad que lo que predeciría una distribución normal, o viceversa. Por otra parte, no podemos rechazar la hipótesis nula de distribución normal en el término remanente. Esto también nos permitió llegar a algunas conclusiones interesantes. Primero, puede significar que las proyecciones climáticas de temperaturas y precipitaciones sean generadas con distribuciones normales. Segundo, la distribución de las expectativas de inflación de los productores parecen distribuirse normalmente. Significa que juzgan poco probable que se llegue a niveles de hiperinflación o deflación altos. Por último, también creemos que las expectativas sobre la tasa de retención se distribuye normalmente. En otras palabras, son pocos los productores que consideran probable que la tasa de retención aumente o se reduzca excesivamente.

En síntesis, sugerimos la utilización del test. Su simple implementación e información desagregada sobre fuente y causa de desvío de normalidad hace del test una herramienta útil para aquellos que vayan a trabajar con datos en panel. Consideramos que el código de Stata no solo debería proveer los p-valores de los tests de hipótesis, sino los coeficientes de kurtosis y simetría. De esta forma, además de saber si la distribución se desvía de la normal, también tendríamos información más detallada sobre la misma.

Universidad de
San Andrés

Referencias

- Aparicio, F. & Estrada, J. (1997). *Empirical Distributions of Stock Returns: Scandinavian Securities Markets, 1990-95*. Working paper, Universidad Carlos III en Madrid, España.
- Breschia, V. & Lema, D. (Junio, 2001). *Modelización de los Aspectos Dinámicos de la Oferta Agropecuaria*. Trabajo presentado en el Taller Internacional “La Modelización en el Sector Agropecuario”, organizado por la Escuela para Graduados “Alberto Soriano”, Facultad de Agronomía, Universidad de Buenos Aires. Buenos Aires.
- Galvao, A. F., Montes-Rojas, G., Sosa-Escudero, W. & Wang, L. (2013). Tests for Skewness and Kurtosis in the One-Way Error Component Model. *Journal of Multivariate Analysis*, 122, 35-52.
- Harri, A., Erdem, C., Coble, K.H. & Knight, T.O. (2009). Crop Yield Distributions: A Reconciliation of Previous Research and Statistical Tests for Normality. *Review of Agricultural Economics*, 31, 1, 163-182
- Risk and Risk Management Strategies in Agriculture: an Overview of the Evidence (27 de Octubre de 2010). *Scottish Government Rural and Environment Analytical Services: Final Report*.
- Wooldridge, J.M. (2009). *Introductory Econometrics: A Modern Approach*. United States of America: South-Western Cengage Learning.
- Xu, F., Hart, C., Smith, D. & Meyers, W.H. (1995). *Estimating the Cost of Revenue Assurance*. Working paper, Center for Agricultural and Rural Development, Iowa State University.

Yu, B., Liu, F. & You, L. (2012). Dynamic Agricultural Supply Response Under Economic Transformation: A Case Study of Henan, China. *American Journal of Agricultural Economics*, 94, 2, 370-376.

Zeckhauser, R. & Thompson, M. (1970). Linear Regression with Non-Normal Error Terms. *The Review of Economics and Statistics*, 52, 3, 280-286.



Universidad de
San Andrés