

Introducción

El texto que a continuación podrán leer, es el resultado de un ejercicio estadístico y de investigación que realicé dentro del Instituto Nacional de Ecología,¹ como parte de un proyecto dirigido por el compilador de este número. De manera precisa, el objetivo de este documento, como su título indica, es hacer un ejercicio econométrico que nos permita, en el marco de la discusión del Cambio Climático (CC), detectar estadísticamente la ocurrencia de eventos poco comunes a la naturaleza típica de una variable que, si bien es simple, hace un buen esfuerzo por capturar la dinámica productiva del sector agrícola en México; por otro lado, habremos de desprender argumentos probabilísticos sobre la sistematicidad de estos eventos que, en adelante, calificaré como extremos. Por esta razón, poco abundaré sobre hechos descriptivos o estadísticos del cambio climático.

Por la relevancia global que el tema del CC ha cobrado desde hace varios años, me parece pertinente realizar ejercicios que expliquen los canales y el impacto del CC en los sectores económicos de forma tal, que se puedan desarrollar las estrategias de mitigación y adaptación pertinentes. La detección y cualificación de este fenómeno climático es hoy día una tarea que se está desarrollando mediante el uso intensivo de modelos. Para los objetivos de nuestro trabajo, podemos decir que una forma intuitiva aunque informal para detectar la ocurrencia de eventos extremos consiste en desarrollar modelos sencillos que muestren, para cualquier punto dentro del espacio muestral y dada una variable dependiente formulada *ad-hoc*, diferencias significativas (y posiblemente sistemáticas) entre el valor observado y el valor pronosticado de la variable. Esta diferencia comúnmente es denominada como término de error o término estocástico (dados ciertos supuestos) y es a partir de este concepto que se deriva lo que conocemos como varianza, que no es otra cosa que la esperanza matemática de estos valores elevados al cuadrado. En modelos poco más complejos, si existe una dinámica propia detrás de la naturaleza de estos eventos (y por ende, detrás de la existencia del término de error), podríamos estar frente a un problema de disturbios

* Estudiante de doctorado en el European University Institute en Florencia, Italia. Comentarios pueden enviarse a Luis.Jaramillo@EUI.eu

¹ Agradezco las facilidades otorgadas por la Dirección Gral. de Investigación en Política y Economía Ambiental (INE-DGIPEA) para realizar este trabajo.

no-esféricos, violando además del principio de homoskedasticidad, el supuesto de no-autocorrelación entre los errores, un requerimiento doble para la matriz de varianza-covarianza. En el formato clásico, la varianza heteroskedástica produce coeficientes estimados caracterizados como insesgados pero ineficientes,² llevándonos a establecer falsamente los argumentos de inferencia estadística; normalmente, regresiones con este tipo de problemas tienden a ser modificadas para su corrección. Para nosotros, los efectos de la heteroskedasticidad mostrada en la matriz de varianza-covarianza más allá de ser un problema, puede significar una alternativa de modelación. Robert F. Engle formuló hace poco más de veinticinco años, la primera versión de lo que hoy conocemos como modelos ARCH (Autoregressive Conditional Heteroskedasticity) y su evolución posterior a GARCH (Generalized Autoregressive Conditional Heteroskedasticity- (Bollerslev, 1986)); estos modelos consideran a la varianza sigue un proceso que debe ser considerado en el desarrollo de modelos econométricos.

Para comprender de mejor manera el problema de la heteroskedasticidad, es útil enfocarlo desde la perspectiva de precisión que se busca exista las estimaciones de un modelo. En los modelos ARCH y GARCH, el elemento clave es la capacidad que tienen para estimar la varianza del término de error, y en su caso, conocer el por qué en ocasiones puede ser alta, no constante o posiblemente agrupada. Desde la perspectiva de nuestro análisis, comportamientos extraños de la varianza del término de error pueden interpretarse como el resultado de eventos extremos en variables altamente sensibles al cambio climático, como son los niveles de temperatura y de precipitación; aún más, en períodos donde la varianza aparentemente obedece a un fenómeno propio, es posible encontrar niveles crecientes de autocorrelación en el término estocástico, soportando fuertemente la hipótesis de que eventos extremos en variables sensibles al CC ocurren cada vez con mayor frecuencia.

Descripción de los datos

Para nuestro análisis, inicialmente utilizamos datos anuales (28 datos) para el período 1980-2007 de las siguientes variables: precipitación promedio anual (PPA), temperatura promedio anual (TPA), superficie sembrada (SSEM), superficie cosechada (SCOS) y valor de la producción (VPM). La Tabla 1 presenta la estadística descriptiva de estas variables.

⁶ Coloquialmente, esta discusión puede plantearse como la elección entre “rapidez o precisión”.

Tabla 1
Descripción de las variables utilizadas

Variable	Siglas	Unidades	Media	Desviación Estándar	Estadístico Chi-Squared
Precipitación Promedio Anual	PPA	Milímetros	68.65	8.13	5.97
Temperatura Promedio Anual	TPA	Grados centígrados	28.31	0.53	0.26
Superficie Sembrada	SSEM	Millones de hectáreas	20.96	1.48	5.24
Superficie Cosechada	SCOS	Millones de hectáreas	18.28	1.71	1.92
Valor de la Producción	VPM	Millones de pesos	92.59	83.77	6.98

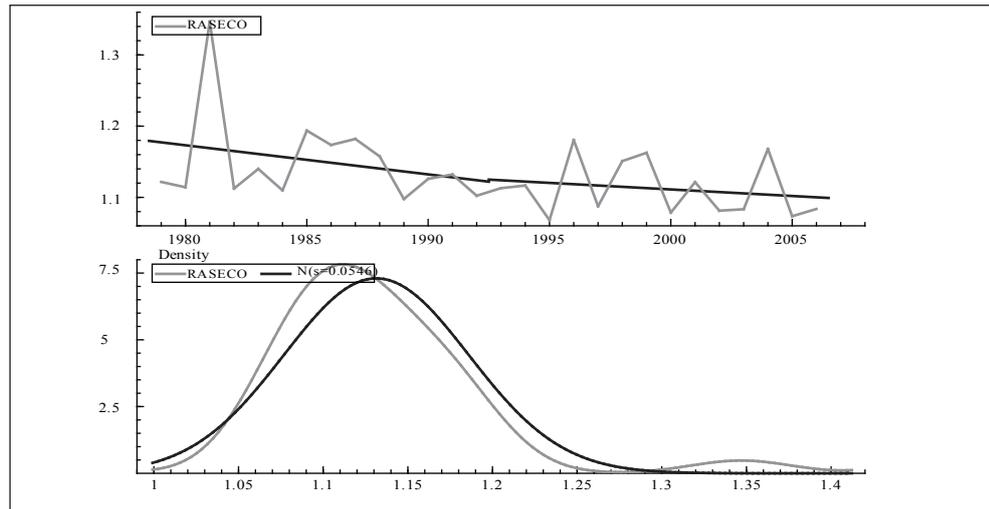
Mediante el análisis estadístico de estas variables, se consideró utilizar la razón superficie sembrada (SSEM) a superficie cosechada (SCOS); esta variable tomó el nombre de RASECO y su descripción estadística se presenta a continuación en la Tabla 2.

Tabla 2

Variable	Siglas	Media	Desviación Estándar	Estadístico Chi Squared (normalidad)
Razón Superficie Sembrada/Superficie Cosechada	RASECO	1.13	0.054	16.78

La razón para utilizar esta variable puede intuirse fácilmente pues cualquier evento extremo observado en variables que anteriormente determinamos como altamente sensibles al CC, tendría forzosamente que reflejarse en la razón RASECO; es decir, comportamientos atípicos en la temperatura o la precipitación, por ejemplo, tendrían un efecto casi inmediato en la superficie cosechada relativo a la superficie sembrada, que es básicamente nuestra variable dependiente.

Gráfica 1



En la gráfica 1 se puede observar que si bien su desviación estándar es particularmente pequeña en relación a la media (demos cuenta de su valor por encima de uno), existe un grado significativo de asimetría positiva en la función de masa de la variable, fortaleciendo la hipótesis, al menos en su planteamiento, de que han existido años en los que la variable RASECO ha mostrado valores claramente atípicos (hipotéticamente generados por cambios significativos en la temperatura o precipitación promedios. Sin embargo, en la parte superior de la gráfica se observa una aparente tendencia hacia el valor que iguala la superficie sembrada y cosechada. De ser así, esto podría implicar que si bien se observan valores extremos (particularmente nos interesan los valores muy por encima de la media), la tendencia en los últimos es hacia valores que no reflejan potencialmente el efecto de cambios drásticos en temperatura y precipitación.

Atendiendo el punto anterior, un procedimiento más o menos sofisticado para analizar posibles tendencias en la serie de datos consiste en una prueba de estacionaridad – particularmente la prueba Dickey Fuller Aumentada (ADF), es decir, realizar una prueba para identificar que la serie podría no ser estacionaria debido a la existencia de una tendencia, o dicho de otra manera, la serie se puede determinar estacionaria siempre y cuando se tome en cuenta el efecto de una tendencia. Los resultados de la prueba ADF se realizaron sobre la serie en sus “niveles” asumiendo una constante y una tendencia, fallando en rechazar la hipótesis nula de raíz unitaria a 1% de significancia estadística,³

³ Este posible fallo en rechazar la hipótesis nula de raíz unitaria puede deberse a la sensibilidad de la prueba a más de un cambio estructural en la serie dado el reducido número de datos disponibles; la existencia de cambios estructurales en la serie soporta de mejor manera, la hipótesis de eventos extremos.

pero rechazando la hipótesis nula a 5%. Estos resultados no nos permiten describir a la serie analizada como “tendencia-estacionaria”, aclarando el punto de una posible tendencia negativa en los datos.

Tabla 3
Prueba de Raíz Unitaria Dickey-Fuller Aumentada de RASECO

ADF Test Statistic	-4.110297	1% Critical Value*	-4.3738
		5% Critical Value	-3.6027
		10% Critical Value	-3.2367

*MacKinnon critical values for rejection of hypothesis of a unit root.

Armados con estos elementos, podemos hacer una estimación para detectar la existencia de efectos ARCH/GARCH en la serie, es decir, si los niveles de heteroskedasticidad sistemáticos en la serie RASECO (desviaciones significativas de la varianza de RASECO respecto de su nivel de largo plazo) pueden, en buena medida, deberse a la existencia de eventos extremos en las variables climáticas, efectos que se reflejan en la relación superficie sembrada/superficie cosechada. A partir de la primera estimación y de acuerdo a los niveles de autocorrelación de los residuos al cuadrado, que existe evidencia parcial de un efecto ARCH debido a que los valores de autocorrelación son positivos y además crecientes, al menos hasta el rezago seis, lo cual podría estar indicando que “la inestabilidad” de la variable RASECO podría ir en aumento.

Tabla 4

	AC	Q-Stat	Prob
1	0.008	0.0018	0.967
2	0.070	0.1619	0.922
3	0.053	0.2558	0.968
4	0.077	0.4642	0.977
5	0.197	1.8849	0.865
6	0.129	2.5196	0.866
7	-0.039	2.5818	0.921
8	-0.025	2.6073	0.957
9	-0.054	2.7379	0.974
10	-0.003	2.7382	0.987
11	0.043	2.8308	0.993
12	-0.056	2.9937	0.996

Es necesario mencionar que si bien existen niveles crecientes y positivos de los parámetros de autocorrelación, estos niveles aparecen como poco significativos. La poca contundencia para argumentar a favor de posibles efectos ARCH/GARCH en la serie RASECO, puede deberse a lo reducido del rango longitudinal de datos (*i.e.* 28 datos). Esta característica puede por un lado estar reflejando la poca significancia de los coeficientes de autocorrelación de los residuos al cuadrado y, por otro, evitando detectar posibles “*clustering* de inestabilidad”, elemento esencial para identificar en el pasado reciente, el efecto de eventos extremos en las variables climáticas sobre la dinámica productiva agrícola.

Simulación de datos bajo eventos extremos

Para hacer frente a este potencial problema, se realizó un ejercicio de simulación de datos donde se tomó cuidadosamente en cuenta, las características de la función de masa de probabilidades de la variable RASECO (es decir, características importantes relativas a la naturaleza de la variable). Esto es particularmente importante considerar pues evidentemente, la variable RASECO “original” es de valor extremo con una significativa asimetría positiva. La parametrización de esta función de masa de probabilidades se realizó utilizando una estimación de *Máxima Verosimilitud* con el programa MATLAB y se obtuvieron los siguientes Momentos estadísticos, con sus respectivos intervalos de confianza.

Tabla 5
Parametrización de la función generadora de momentos de la variable RASECO

	Parámetro	Intervalos de Confianza	
		Inferior	Superior
Scale	0.10471	-0.0744	0.2839
Shape	15.15	9.14	21.16

La simulación de 50 datos aleatorios generada tomando como referencia los Momentos estadísticos antes señalados, se observa de la siguiente manera:

Los resultados de esta simulación de datos que denominaremos SRASECO, no tienen problema alguno de no-estacionaridad (ver tabla 6) pues rechazan la hipótesis nula de raíz unitaria a cualquier nivel de significancia estadística y además, conservan las propiedades asimétricas de la distribución original. Esto es particularmente importante pues son precisamente estos datos ubicados en la cola derecha de la distribución, los que contienen información acerca de la existencia de los efectos de eventos climáticos extremos en la variable RASECO.

Gráfica 2
Función de masa de los datos simulados para RASECO

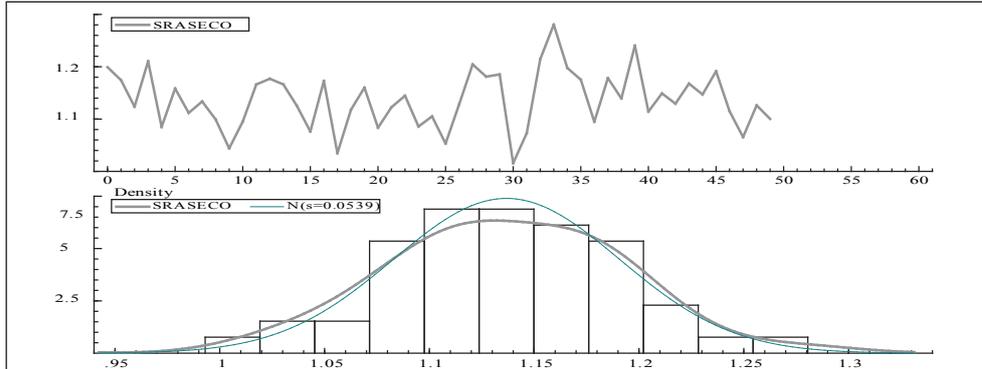


Tabla 6
Prueba de Raíz Unitaria Dickey-Fuller Aumentada de SRASECO

ADF Test Statistic	-4.649303	1% Critical Value*	-3.5713
		5% Critical Value	-2.9228
		10% Critical Value	-2.599

*MacKinnon critical values for rejection of hypothesis of a unit root.

Resultados

No está por demás comentar que la simulación tan brevemente presentada, requirió de un cuidadoso proceso estadístico cuyos detalles para efectos didácticos, no se antojan como necesarios; sin embargo, es pertinente mencionar que procesos de simulación similares⁴ en procedimiento pero con alcances mayores, pueden ser generados para enriquecer el análisis. Los resultados presentados en el cuadro 1 muestran la ventaja de trabajar con una mayor extensión longitudinal de datos que, de no ser por la simulación, hubiera sido prácticamente imposible disponer o en su defecto, podría disponerse pero con un muy bajo nivel de confiabilidad. Estos resultados, incluso en su versión más simple, son por demás interesantes.

Primero, la significancia estadística de los coeficientes estimados muestran potencialmente, los efectos de eventos extremos contenidos en la variable RASECO, una variable particularmente sensible al cambio climático; es decir, el corte de datos refleja la existencia de desviaciones significativas de

⁴ Mayores detalles concernientes al proceso de simulación están disponibles para los interesados.

la varianza de su nivel de largo plazo (*i.e.*, refleja un comportamiento heteroskedástico). Aún más, la detección de este fenómeno conocido como efecto GARCH cobra mayor relevancia si consideramos que la existencia de estas desviaciones no podría “mapearse aleatoriamente” a lo largo de la serie, sino que existe un nivel de autocorrelación en ellas. Esto último puede significar que la observación reciente de este efecto tiene la capacidad de pronosticar réplicas de este comportamiento en el futuro cercano. Por otro lado, dado que el valor de los coeficientes es cercano a uno, se desprende que en estos efectos GARCH existe un proceso denominado de “reversión a la media” en extremo lento (es decir, de regreso al valor de largo plazo), implicando básicamente dos cosas: *a*) en el corto plazo, no aparece como plausible dejar de observar estos efectos y, *b*) que las desviaciones de la varianza respecto de su media, cuando se observan, son significativamente grandes. La combinación de estos dos efectos estadísticamente detectados, puede resultar altamente peligroso dentro del marco de análisis de los efectos del cambio climático en la agricultura.

Cuadro 1
Resultados del Modelo GARCH(1,1)

Dependent Variable: SRASECO				
Method: ML - GARCH				
Convergence achieved after 49 iterations				
Bollerslev-Wooldrige robust standard errors & covariance				
	Variance Equation			
C	0.000798	0.000420	1.899405	0.0575
ARCH(1)	0.671038	0.260862	2.572384	0.0101
GARCH(1)	0.151293	0.198049	0.763918	0.1249
R-squared	0.181978	Mean dependent var		1.138621
S.E. of regression	0.056929	Akaike info criterion		-2.789925
Sum squared resid	0.097229	Schwarz criterion		-2.406026
Log likelihood	63.40354	F-statistic		0.834229

Bibliografía

- B. Nelson Daniel (1991), Conditional Heteroskedasticity in Asset Returns: A New Approach, *Econometrica*, vol. 59, núm. 2, marzo 1991, 347-370.
- Bollerslev Tim (1986), Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, vol. 31, 307-327.
- Hall Peter y Qiwei Yao (2003), Inference in ARCH and GARCH Models with Heavy-Tailed Errors, *Econometrica*, vol. 71, núm. 1, enero 2003, 285-317.