AgEcon Search Appendix II to: "Crop Insurance Savings Accounts: A Viable Alternative to Crop Insurance?" to appear in

Applied Economic Perspectives and Policy in 2014

Author: Octavio A. Ramirez

Given the procedures to be followed in this research, time series of price and yield realizations that are representative of what farmers might face in future years are needed to evaluate the feasibility of the proposed CISA. Reliable parametric estimates of future price and yield distributions are required to generate those realizations and sufficiently long historical price and yield time series are necessary in order to estimate those distributions. While long time series are available for most major commodity prices, multi-decade farm-level yield records are not as common. Fortunately, the University of Illinois Endowment Farms project has been collecting such records from 26 different "representative" corn producers during the last 50 years. Therefore, the "test-of-concept" analyses presented in this article are conducted for the specific case of corn producers in the State of Illinois.

Price and Yield Distribution Models

In addition to having access to suitable data, a key to obtaining realistic estimates of the price and yield distributions of interest is to use flexible probability density function (pdf) models that can accommodate a wide range of mean-variance-skewness-kurtosis combinations. One such density function is the Inverse Hyperbolic Sine (IHS), which was first utilized for yield modeling and simulation by Ramirez (1997). Subsequent applications of this model involving both yield and price distributions include Ramirez and Somarriba (2000), Ramirez, Misra and Field (2003), and Ramirez, McDonald and Carpio (2010).

In addition to its flexibility, the IHS distribution model is appealing because each of its first four statistical moments can be independently controlled by a parameter or a parametric function of some exogenous variable(s). Specifically, for both the price and yield distributions, the mean is specified as a linear function of time ($B_1 + B_2 t,\ t = 1,2,\dots,T$) while the variance, skewness and kurtosis are controlled by constant parameters (B_3, B_4, B_5, respectively). In the single variable case, the IHS density is then given by:

(1) $IHS(Y_t) = G_t(2\pi)^{-\frac{1}{2}}\exp(-0.5H_t^2)$, where

$$G_t = [B_3^2(1 + R_t^2)/J]^{-\frac{1}{2}},$$

$$J = [\exp(B_4^2) - 1][\exp(B_4^2)\cosh(-2B_4B_5) + 1]/(2B_4^2),$$

$$R_t = J^{\frac{1}{2}}B_4(Y_t - B_1 - B_2t)/B_3 + F,$$

$$F = \exp(0.5B_4^2)\sinh(B_4B_5),$$

$$H_t = \ln\,(R_t + (1 + R_t^2)^{\frac{1}{2}}/B_4) - B_5.$$

As Ramirez, Misra and Field (2003) point out, as B_4 and B_5 approach zero, this pdf becomes a normal density with mean $B_1 + B_2t$ and variance B_3^2, which facilitates a test for whether or not prices and yields are normally distributed. In addition, if $B_4 \neq 0$ but $B_5 = 0$, the density is kurtotic but symmetric, while a negative (positive) B_5 induces negative (positive) skewness into the distribution. Specifically, the skewness (S) and kurtosis (K) measures of this pdf are given by:

(2) $S = \frac{1}{4}W^{\frac{1}{2}}(W-1)^2[W(W+2)\sinh(3Q) + 3\sinh(Q)]/(JB_4{}^2)^{1.5}$, and

(3) $K = \{(W-1)^2[W^2(W^4+2W^3+3W^2-3)\cosh(4Q)+4W^2(W+2)\cosh(2Q)+3(2W+1)]/8J^2B_4^4\}-3$, where

(4) $W = \exp(B_4^2)$ and $Q = -B_4B_5$.

In short, the IHS model allows for a wide range of skewness-kurtosis combinations (according to the two equations above which only depend on B_4 and B_5) while its mean and variance are determined by $B_1 + B_2t$ and B_3 only. In addition, Ramirez, Misra and Nelson (2003) show how the IHS density (equation 1) can be modified to allow for autocorrelation. Specifically, all is needed is to let $R_t = \left(J^{\frac{1}{2}}B_4P_t(Y_t - B_1 - B_2t)/B_3\right) + F$ where P_t is the t^{th} row of a T by T transformation matrix P such that $P'P = \Psi^{-1}$ and Ψ is the error term correlation matrix (Judge et al. 1985). Using standard procedures, the concentrated log-likelihood function needed for estimating the parameters of this model can be derived from equation (1):

(5) $\sum_{t=1}^{T} \ln(G_t) - 0.5\sum_{t=1}^{T} H_t^2$

The above function is then maximized in order to obtain estimates for the parameters of a price distribution model with a time-varying mean, constant variance, skewness and kurtosis coefficients, and a suitable autocorrelation process. Maximum likelihood estimation is accomplished using the CML procedure of Gauss 9. The data utilized includes the real (inflation-adjusted[2]) corn prices received by Illinois farmers during the last 70 years (USDA, National Agricultural Statistics Service 2011). As customary, the price series is first tested and confirmed to be stationary according to both the Dickey-Fuller and the Phillips-Peron tests.

The maximum-likelihood parameter estimates and related statistics for this first model are presented in table 1. First note that real prices have been decreasing over time at a rate of 3.22 cents/year, putting them at a predicted average of $4.085/bushel in 2011. The estimate for the standard deviation of the price distribution stands at $0.618/bushel. A White test is conducted to make sure that the model's variance is constant, i.e. that price variability has not been changing over time. A test statistic of 3.37 does not allow for the rejection of the null hypothesis of homoscedasticity (p-value= 0.185).

[Insert Table 1 about Here]

The maximum value of the concentrated log-likelihood function corresponding to the non-normal price model is -60.37 versus -64.92 for the analogous normal model where B_4 and B_5 are set to zero. As a result, the likelihood ratio test statistic (Ramirez, Misra and Field 2003) easily allows for rejection of the null hypothesis of normality (p-value=0.01). That is, since both B_4 and B_5 are positive, the distribution of corn prices received by farmers in the state of Illinois is in fact positively kurtotic and significantly right-skewed. Finally it is evident that, over time, prices follow a second order autoregressive process as both parameters in this process (B_6 and B_7 in the transformation matrix P) are highly significant while the Box-Pierce test cannot reject the null hypothesis that the transformed model residuals $\{P_t(Y_t - B_1 - B_2 t)\}$ are independently distributed (p-value=0.978). As described in the next section, this model can be used to obtain draws from the current and future price distributions for the purposes of the CISA analyses.

Farm-level yield models are also estimated using the previously described procedures, assuming that there is no autocorrelation. The data in this case is obtained from the University of Illinois Endowment Farms project. Specifically, their ten farms with the largest sample sizes (40 to 45 years) are selected for inclusion in the analyses. The maximum-likelihood parameter estimates and related statistics for these 10 yield distribution models are presented in table 2.

[Insert Table 2 about Here]

First note that all yields are increasing over time, with the rate of increase averaging about 1.4 bushels/acre per year. The predicted yields for 2011, presented in the first row of the table, average a little over 170 bushels/acre versus about 115 bushels/acre in the early 1970s. The standard deviation parameters of the yield distributions range from 18 to 30 bushels/acre and, as with prices, the White tests statistics (also reported in table 2) suggest that yield variability has generally remained constant over the last 40 years. The null hypothesis of yield normality is strongly rejected (p-value<0.025) in four cases, rejected (p-value<0.10) in two cases, and cannot be rejected in the remaining four. In contrast to prices, the prevailing negativity in the B_5 estimates suggests that the yield distributions tend to be left-skewed. Two of the non-rejection instances might be explained by the fact that, in both cases, observations were missing for the year 1983 which was characterized by extremely low yields in most other farms. In the other two, it appears that somehow farmers managed to avoid an extremely low yield event during the observation period, which is needed to trigger rejection.

Price and Yield Simulation

The process of simulating draws from an estimated IHS pdf is simplified by the fact that the IHS random variable is actually defined as a function of a normal (Ramirez 1997). Specifically, if Z_t is a standard normal, then:

(6) $$IHS_t = mean_t\{sig(sinh(\theta(Z_t+\mu)) - F)/(\theta J^{1/2})\},$$

where F and J are as defined in equation (3) and, in reference to the models in the previous section, $mean_t = B_1 + B_2 t$, $sig = B_3$, $\theta = B_4$, and $\mu = B_5$. Thus, once an IHS distribution model parameters have been estimated, random draws from the implied distribution can be easily obtained on the basis of standard normal draws. In addition, contemporaneously correlated draws from several (S) IHS variables can be generated by simply correlating the (1 by S) Z_t vectors used to generate them by the Cholesky decomposition of the desired (S by S) correlation matrix (Ramirez 1997). Finally, when the estimated IHS model involves autocorrelation, any T draws can be made to follow that process by multiplying a vector of IHS errors ($\{IHS_t - mean_t\} = \{sig(sinh(\theta(Z_t + \mu)) - F)/(\theta J^{1/2})\}, t = 1, \ldots, T$) by the Cholesky decomposition of the appropriate correlation matrix $\Psi = (\mathrm{P}'\mathrm{P})^{-1}$ and then adding back the systematic component of the model ($mean_t$).

The above procedures are used in conjunction with the estimated model parameters to simulate random realizations of prices and yields to be experienced by NF=10,000 hypothetical corn farms in the State of Illinois. It is assumed that the population of 10,000 farms is equally divided into 10 groups, each of which is characterized by one of the 10 yield distributions models detailed in table 3 (six non-normal and four normal). Forty-five future years of random yields are simulated for each farm assuming correlations of 0.65 across all yield distributions. In addition, 40 years of future state-wide price realizations are simulated assuming correlations of -0.45 with each of the 10,000 sets of yield draws. The 0.65 yield-yield correlation is selected on the basis of the average of the 45 sample correlation coefficients observed across the 10 farm-level yield series underlying the analyses. The -0.45 yield-price correlation is based on the average of the 10 sample correlation coefficients observed between the 10 yield series and the state-wide price data during the period those yields were observed.

References

Judge, G.G., Griffiths, W.E., Carter Hill, R., Lutkepohl, H., and Tsoung-Chao Lee. 1985. *The Theory and Practice of Econometrics*. 2nd ed. New York: John Wiley & Sons, Inc.

Ramirez, O.A. 1997. “Estimation and Use of a Multivariate Parametric Model for Simulating, Heteroskedastic, Correlated, Nonnormal Random Variables: The Case of Corn Belt Corn, Soybean and Wheat Yields.” *American Journal of Agricultural Economics* 79(1):191–205.

Ramirez, O.A., S. Misra, and J. Field. 2003. “Crop-yield Distribution Revisited.” *American Journal of Agricultural Economics* 85(1): 108-120.

Ramirez, O.A., S.K. Misra, and J. Nelson. 2003. “Efficient Estimation of Agricultural Time Series Models with Non-normal Dependent Variables.” *American Journal of Agricultural Economics* 85(4): 1029–1040.

Ramirez, O.A., T.U. McDonald, and C.E. Carpio. 2010. “A Flexible Parametric Family for the Modeling and Simulation of Yield Distributions.” *Journal of Agricultural and Applied Economics* 42(2): 1-17.

Ramirez, O.A., and E. Somarriba. 2000. “Risk and Returns of Diversified Cropping Systems under Non-normal, Cross and Autocorrelated Commodity Price Structures. *Journal of Agricultural and Resource Economics* 25(2): 653-668.

Table 1. Maximum-Likelihood Parameter Estimates and Related Statistics for the Non-Normal Price Distribution Model

	P.E.	S.E.E	T.V.	P.V
B_1	6.3412	0.2144	29.5815	0.0000
B_2	-0.0322	0.0052	6.2110	0.0000
B_3	0.6179	0.0745	8.2948	0.0000
B_4	0.3229	NA*	NA*	0.0106
B_5	20.0914	NA*	NA*	0.0106
B_6	0.7605	0.1091	6.9694	0.0000
B_7	-0.3974	0.1228	3.2354	0.0010

Notes: P.E., S.E.E, T.V., and P.V. stand for parameter estimate, standard error estimate, t-value and p-value respectively. The significance (p-value) of the non-normality parameters (B_4 and B_5) is ascertained through a likelihood ratio test. B_6 and B_7 are the first- and second-order autoregressive parameters.

Table 2. Maximum-Likelihood Parameter Estimates of Yield Distribution Model

	Farm 1		Farm 2		Farm 3		Farm 4		Farm 5	
	N	**NN**	**N**	**NN**	**N**	**NN**	**N**	**NN**	**N**	**NN**
Mean	182.39	193.58	162.14	161.47	179.00	183.10	174.77	173.44	163.47	163.47
B1	100.69	92.60	85.39	84.98	93.38	89.89	88.06	89.57	99.78	99.78
B2	1.542	1.905	1.448	1.443	1.616	1.759	1.636	1.583	1.202	1.201
B3	20.926	21.236	18.005	20.931	22.695	23.477	18.327	25.820	20.301	20.301
B4	0.000	0.914	0.000	0.808	0.000	0.722	0.000	1.258	0.000	0.000
B5	0.000	-0.436	0.000	-0.683	0.000	-0.787	0.000	-0.041	0.000	0.000
Skew	0.000	-2.405	0.000	-2.251	0.000	-1.808	0.000	-1.147	0.000	0.000
Kurt	0.000	28.144	0.000	17.222	0.000	10.416	0.000	306.377	0.000	0.000
White	2.318	2.227	3.741	3.801	2.635	2.456	2.228	2.409	4.831	4.831
-2MV	392.48	375.48	379.24	377.05	390.53	381.57	372.15	363.85	398.67	398.67
LRTS		16.994		2.191		8.965		8.297		0.000

	Farm 6		Farm 7		Farm 8		Farm 9		Farm 10	
	N	**NN**	**N**	**NN**	**N**	**NN**	**N**	**NN**	**N**	**NN**
Mean	168.15	181.43	165.97	169.30	186.07	188.71	165.95	171.69	136.04	140.88
B1	114.63	103.19	89.29	86.20	121.66	118.99	128.67	123.16	84.49	80.77
B2	1.010	1.474	1.447	1.568	1.215	1.315	0.704	0.916	0.973	1.134
B3	25.492	27.087	27.705	29.943	21.424	23.122	24.481	26.618	25.454	25.717
B4	0.000	0.418	0.000	0.735	0.000	0.518	0.000	0.725	0.001	0.273
B5	0.000	-15.000	0.000	-0.775	0.000	-9.451	0.000	-0.723	0.000	-15.000
Skew	0.000	-1.394	0.000	-1.876	0.000	-1.832	0.000	-1.717	0.000	-0.856
Kurt	0.000	3.642	0.000	11.257	0.000	6.509	0.000	10.002	0.000	1.330
White	4.846	4.166	2.579	2.252	4.539	4.558	1.494	1.642	3.290	3.493
-2MV	391.21	385.31	398.21	392.61	358.68	348.59	369.35	365.73	409.71	405.69
LRTS		5.907		5.598		10.089		3.621		4.024

Notes: N and NN stand for normal and non-normal model, respectively. Skew and Kurt are the standard measures of kurtosis and skewness. White is the White test statistic which, under the null hypothesis of homoscedasticity, is distributed as a $\chi^2_{(2)}$ random variable. -2MV is minus two times the maximum value of the log likelihood function and LRTS is the resulting likelihood ratio test statistic which, under the null hypothesis of normality, is also distributed as a $\chi^2_{(2)}$ random variable.