Forecasting World Crop Yields as Probability Distributions

John N. (Jake) Ferris
Professor Emeritus
Department of Agricultural Economics
Michigan State University

Contributed paper prepared for presentation at the
International Association of Agricultural Economists Conference
Gold Coast, Australia
August 12-18, 2006

The traditional approach to forecasting crop yields in large econometric models of agriculture is to assume normal weather and project crop yields as linear extensions of past trends. Recognizing that departures of yields from trends are the major sources of risk in the year ahead and beyond, agricultural economists have made the case for probability forecasting for many years (Teigen and Bell, 1978; Nelson, 1979; Ikerd, 1979). Ferris (1989) illustrated the procedure to introduce stochastic yields in an econometric model to forecast crop prices three years ahead and enumerated the process in his textbook (Ferris, 1998, 2005). In recent years, Richardson (2000) has provided leadership in simulating risk from variations in crop yields and other sources as applied to farm management decisions. Ray along with Richardson (Ray, Richardson, et al., 1998) has undertaken an extensive effort to generate long term agricultural forecasts employing probability distributions of crop yields and exports.

After this paper was submitted to IAAE, two studies were published which applied stochastic modeling to policy analysis, both dealing with problems of asymmetry. Both cited the weakness in point forecasts, one dealing with projections under WTO restrictions (Westhoff, et al., 2005) and the other with regard to government payments for 2007 to 2015 under the current U.S. 2002 Farm Act (U.S. Department of Agriculture, 2006).

Opportunities Have Been There

Expanding computer capacity has increasingly enabled agricultural economists to generate probability distributions of variables in econometric/simulation models. Obviously, crop yields are the most attractive candidates. Long term historic data on crop yields are widely available, so estimating the variability coefficients is not difficult. Variability may be changing, which must be assessed in order to extrapolate into the future. In this case, the analyst may want to test whether departures from rising crop yield trends are more stable in percentage terms than absolute terms. The analyst may also face the question of whether to eliminate "outliers" such as departures from trend due to extreme weather which is considered extremely unlikely in the future, such as a flood that happens only once in 500 years. These problems are relatively easy to solve.

Challenges

Why only limited attention has been give to stochastic forecasting by modelers is clear when one considers the challenges. Beyond the enormous capacity requirements for computers to handle stochastics, analysts for large econometric models face the following major frontiers: (1) incorporating crop yield correlations among crops and (2) generating non-normal distributions. This paper addresses these challenges with an empirical example employing the software program, EViews 5 (Quantitative Micro Software, 2004). Quantitative Micro Software provided technical support in programming for this exercise.

Selected Crops

AGMOD, an econometric/simulation model of United States agriculture developed at Michigan State University, focuses on major field crops and livestock incorporating yields on thirteen selected crops and regions in the world. Crop yields per acre or per hectare from a data base of the Foreign Agricultural Service (PSD Online) of the U.S. Department of Agriculture (U.S. Department of Agriculture, 2005) were divided by double exponential smoothing of the data to estimate departures from trends in a ratio form. The codes and definitions for these ratios are as follows:

RYCN	US corn
RYCGOCN	US coarse grains other than corn

RYSB	US soybeans
RYWH	US wheat
RYCGX	Coarse grains in major exporting nations of Canada, Argentina, and Australia
RYWHX	Wheat in major exporting nations of Canada, Argentina, and Australia
RYCGEU	Coarse grains in the 15 nations of the European Union (EU15)
RYWHEU	Wheat in the EU15
RYOLSEU	Oilseeds in the EU15
RYSBAB	Soybeans in Argentina and Brazil
RYCGOXE	Coarse grains in nations other than the US, Canada, Argentina, Australia and the EU15
RYWHOXE	Wheat in nations other than the US, Canada, Argentina, Australia and the EU15
RYOLSO	Oilseeds in nations other than soybeans in the US, Argentina and Brazil, and oilseeds in the EU15

Yield Correlations

A correlation matrix was generated for these thirteen crop variables as presented in Table1. Most of the yield data were for the period of 1960 to 2004 with exceptions as noted. As expected, the positive correlations are strongest for the regions that are overlapping such as (1) the production areas for US corn, other coarse grains and soybeans, (2) coarse grains and wheat in Canada, Argentina, and Australia, (3) coarse grains, wheat and oilseeds in EU15, and (4) coarse grains and wheat in the rest of the world.

Also of interest is the fact that some of the correlations are negative, such as between RYSBAB and RYCGOXE, RYSBAB and RYOLSO and RYCN and RYCGOXE. These negative correlations provide offsetting impacts of weather on world crop yields contributing to crop production stability.

To correct for the strong positive correlations, Ordinary Least Squares (OLS) procedures were applied to forecast RYSB and RYCGOCN as functions of RYCN. Similarly, equations were estimated for RYWHX as a function of RYCGX; RYWHEU and RYOLSEU as functions of RYCGEU; and RYWHOXE as a function of RYCGOXE.

Normality Tests

Standard deviations were estimated for each of the yield variables, and Jarque-Bera (Bera, Jarque, 1980) normality criteria were also applied to the distributions (Table 2). Under the Jarque-Bera test, a normal distribution has a skewness value of zero and a kurtosis value of three. Negative values on skewness mean longer tails to the low side than to the high side. Values of kurtosis over three indicate thicker tails than for a normal distribution. Note in Table 2 that crop yields tend to be skewed to the low side.

For RYCN, the standard deviation was .095 or 9.5 percent of trend. The distribution was skewed to the low side and with thick tails. The Jarque-Bera test suggests that there is little likelihood that RYCN is distributed normally. In other words, yields are lower in years of unfavorable weather than higher in years of favorable weather. The Jarque-Bera criteria for RYSB indicated a similar characteristic but less pronounced. By generating RYSB as a function of RYCN, the residual term in the equation displayed normality. While the normality criteria for RYCGOCN were positive, the residual from the regression on RYCN exhibited stronger normality features.

RYCGX approached normality while RYWHX did not. By regressing RYWHX on RYCGX, the residual term improved on the normality criteria.

Like RYCN, RYCGEU was skewed to the low side with thick tails. RYWHEU was also not normal but tended to be skewed to the high side. Regressing RYWHEU on RYCGEU generated residuals closer to normal but not to a strong position. Regressing RYOLSEU on RYCGEU helped to explain oilseed yields, but the residuals did not represent an improvement in terms of normality as compared to that of RYOLSEU itself. In any case, both the variability of RYOLSEU and the residuals from the regression on RYCGEU were acceptably normal at .92 and .62 probability levels respectively.

RYSBAB over the 1965 to 2004 period was normally distributed based on Jarque-Bera. The period of analysis was shortened for Argentina and Brazil because of the rapidly increasing production level for those countries. An even shorter period was applied to RYOLSO because of the much higher level of production since 1987 than before.

Because of the dispersion geographically of RYCGOXE, the variability around trend approached normality over the 1960 to 2004 crop years as did RYWHOXE (Table 2). The residuals from the regression of RYWHOXE on RYCGOXE displayed a stronger position on normality than did the direct test on RYWHOXE.

RYOLSO was such an aggregate variable that it was not highly correlated with other crop variables, and it exhibited reasonable normality at a .721 probability coefficient under Jarque-Bera. This is another example that deviations of crop yields from trends tend toward normalcy if broad geographic areas are involved.

Yield Simulations for 2006

With the above information relative to correlations and normality criteria, a model was constructed to simulate yield forecasts in the year 2006 as probability distributions. For each yield, a single valued projection was established for 2006 based primarily on past trends. For yields correlated with and regressed on other yields, the equation for the ratio to trends was a function of the other yields and a random number generator which returned a normal distribution with a mean of zero and a standard deviation connected to the residuals of the equation. For crops for which yield deviations were classified as normal, the equation for deviation from trends was simply a function of the standard deviation times the random number generator for a normal distribution.

To illustrate this sector of the model are the lines for US soybeans and wheat:

RYSBP = .469 + .549*RYCNP+ .0557*NRND

RYWHP = 1.000 + .0695*NRND

Where: RYSBP = Predicted ratio of US soybean yields to trends

RYCNP = Predicted ratio of US corn yields to trends

RYWHP = Predicted ratio of US wheat yields to trends

NRND = Random number generator for a normal distribution

The next sector of the model transformed the ratio distributions into the yield distributions around the projected levels. For US corn, soybeans and wheat, the lines of the model are as follows:

YCNP = RYCNP*YCN

YSBP = RYSBP*YSB

YWHP = RYWHP*YWH

Where: YCN = Single value projection of US corn yields to 2006

YSB = Single value projection of US soybean yields to 2006

YWH = Single value projection of US wheat yields to 2006

However, a program is necessary to run a stochastic version of this model. This program first has to generate non-normal distributions for the two problem distributions which deviate from normal: US corn yields and EU15 coarse grain yields. To do this, the ratios of actual to trend yields for 1960 to 2004 for these crops were drawn randomly. These draws then provided the inputs for the model for RYCNP and RYCGEU to generate the probability forecasts for 2006.

The next question is how many times should you run the model? The arbitrary decision in this case was 1000. Fortunately, the power of desktop computers will allow us to explore even more runs. To generate forecasts with risk components into the next 5 or 10 years, many more iterations will be needed.

Results for US Corn, Soybeans and Wheat Yields

The results for the one year of 2006 are illustrated for YCNP, YSBP AND YWHP in Figures 1, 2, and 3. On YCNP, the most frequent yield forecast for 2006 was around 150 bushels per acre; but because of the skewed distribution, the mean is 143 bushels and the median is 145 bushels per acre. The probability of a short crop of 130 bushels or less is about 18 percent and about the same probability for a large crop exceeding 155 bushels per acre.

Note the differences in Figures 1 to 3 on how the yield distributions were calculated. In Figure 1, the random draws of deviations from trends on US corn were from actual observations for 1960 to 2004 because the distribution was non-normal. While deviations from trends on soybeans were also skewed to the low side, the distribution could be simulated by a regression on corn yield deviations in combination with random draws from a normal distribution with the standard deviation in the residuals as the parameter for the normal distribution. Deviations from trends on wheat yields have approached normality; so to generate the forecast for 2006, the projected mean yield was multiplied by a random number generator for a normal distribution with the standard deviation of .069 calculated from RYWH for 1960 to 2004.

The yields on corn demonstrated the lack of normality as expected with a standard deviation of 13.4 bushels per acre or .094 as a ratio to the mean, close to .095 as a ratio to trend yield. Skewness was -1.036 in the simulation as compared to -1.048 in the 1960 to 2004 period. Kurtosis was 3.769 as compared to 3.842 in the actual deviations in 1960 to 2004.

On US soybeans, the standard deviation was 3.1 bushels per acre or .075 as a ratio to the trend yield, close to the .076 in the data base. Skewness was -.385, also close to -.326 in the data base with kurtosis at 3.394, also close to 3.456 in the data base. On US wheat was a standard deviation of .069 as a ratio to trend yields (2.9 bushels per acre) compared to a standard deviation of .070 in the data base. Skewness and kurtosis remained in the general bounds for a normal distribution.

Implications

The implications of this analysis for 2006 can be extended and applied in a more general way. Integrated into an econometric model of US agriculture, these distributions provide the basis for forecasting domestic production, exports, imports, prices received by farmers, farm incomes, government expenditures, food prices, etc. in a dimension which allows for risk assessment. This information can enter other models for decision makers with additional detail to fit the situation. Single value projection sets which serve as the most likely scenario can thereby be substantially broadened for a much richer view of the future.

While the focus in this application of probability yield generation is on 2006, projecting with econometric models stochastically beyond one year ahead must be carefully programmed. Each iteration in the solution of the model must cover the entire projection period. The reason is that the random draws from the crop yield distributions for the first year of the projection period will affect agricultural variables the next year; the results for the second year will affect the third year; etc. For example, if the draw in the first year results in short crops and high prices, farmers will tend to increase acreages in the second year resulting in lower prices and so on. This, of course, adds progressively to number of iterations required to simulate accurate distributions.

Stochastic modeling is especially important in periods of low world stock levels on grain and oilseeds in planning for food security and helping to avoid spikes in food prices. On the other hand, large crop carryovers weigh heavily on government expenditures on farm programs and increase demands for careful forecasts for budgeting purposes. As cited by Westhoff (Westhoff, et al., 2005) and the USDA (U.S. Department of Agriculture, 2006), point projections can be misleading in assessing future costs of domestic farm programs and WTO proposals. Stochastically derived projections pick up the costs and other items of interest missed by scenarios assuming normal crop yields.

In addition to risk assessments for policy analysis, large econometric models solved stochastically can provide useful inputs into type of farming models for decision making. As in the past, interest will remain for “the most likely scenario” assuming normal weather. Even so, computer capacity and software technology are available to make probability forecasting more routine in the future.

References

Bera, A., Jarque, C., 1980. Efficient Tests for Normality, Heteroscedasticity and Serial Independence of Regression Residuals, Economics Letters, 6.

Ferris, J., 1989. Generation of Probability Distributions on Farm Price and Supply/Utilization Balance on Corn, Soybeans, and Wheat for the Crops of 1989-91, Paper presented at the NCR-134 Conference on Applied Commodity Price Analysis, Forecasting and Market Risk Management., CME, April 20-21, 1989.

Ferris, J., 1988, 2005. Agricultural Prices and Commodity Market Analysis. WCB McGraw-Hill, 1998. Michigan State University Press, 2005.

Ikerd, J., 1979. Using Probability Estimates in Outlook Work. Paper for the Southern Regional Outlook Conference, Atlanta, GA.

Nelson, G., 1979. The Case for Probabilistic Outlook Information for Agricultural Commodities. Paper for Great Plains and Western States Outlook Conference. Moscow, ID.

Quantitative Micro Software, LLC, 2004. EViews 5. Irvine, CA.

Ray, D., Richardson, J., Ugarte, D, Tiller, K.,1998. Estimating Price Variability in Agriculture: Implications for Decision Makers. Journal of Agricultural and Applied Economics, 30, 1.

Richardson, J., Klose, S., Gray, A., 2000. An Applied Procedure for Estimating and Simulating Multivariate Empirical (MVE) Probability Distributions In Farm-Level Risk Assessment and Policy Analysis. Journal of Agricultural and Applied Economics, 32, 2.

Teigen, L, Bell, T., 1978. Confidence Intervals for Corn Price and Utilization Forecasts. Agricultural Economics Research, U.S. Department of Agriculture, 30, 1.

U.S. Department of Agriculture, Foreign Agriculture Service, 2005. Production, Supply and Distribution (PSD Online), http://www.fas.usda.gov/psd/.

U.S. Department of Agriculture, World Agricultural Outlook Board, 2006. USDA Agricultural Baseline Projections to 2015.

Westhoff, P., Brown, S., Hart, C., 2005. When Point Estimates Miss the Point: Stochastic Modeling of WTO Restrictions. FAPRI Policy Working Paper #01-05. Food and Agricultural Policy Research Institute. University of Missouri/Iowa State University.

Table 1. Correlation Matrix of Ratios of Actual to Trend Yields on Thirteen Selected Crops and Regions of the World, 1960 to 2004[a]

	RYCN	RYCGOCN	RYSB	RYWH	RYCGX	RYWHX	RYCGEU	RYWHEU	RYOLSEU	RYSBAB	RYCGOXE	RYWHOXE	RYOLSO
RYCN	X	0.690	0.685	0.272	0.203	0.040	0.066	0.059	0.226	-0.164	-0.230	-0.042	0.315
RYCGOCN		X	0.770	0.303	0.442	0.466	0.152	0.118	0.036	0.024	-0.043	-0.108	0.245
RYSB			X	0.065	0.154	0.061	0.055	-0.106	0.079	0.015	-0.148	-0.062	0.066
RYWH				X	0.354	0.235	0.057	0.104	-0.022	-0.004	-0.100	0.151	0.190
RYCGX					X	0.609	0.014	-0.094	0.339	0.174	-0.045	0.058	0.141
RYWHX						X	-0.059	-0.098	-0.156	-0.072	0.118	0.111	-0.063
RYCGEU							X	0.667	0.535	0.276	0.061	-0.010	0.163
RYWHEU								X	0.332	-0.005	0.243	0.074	0.356
RYOLSEU									X	0.020	0.198	0.135	0.187
RYSBAB										X	-0.290	0.084	-0.243
RYCGOXE											X	0.628	-0.129
RYWHOXE												X	-0.179
RYOLSO													X

[a] For RYCGOCN and RYSBAB, the years are 1965 to 2004.
For RYOLSEU and RYOLSO, the years are 1980 to 2004.

Table 2. Variability and Normality Measurements on Ratios of Actual to Trend Yields on Thirteen Selected Crops and Regions of the World, 1960 to 2004[a]

Crop	Standard Deviation (DV)	Skewness	Kurtosis	Probability of Normality from Jarque-Bera
RYCN	0.095	-1.036	3.769	0.010
RYCGOCN	0.085	-0.044	2.531	0.827
RYSB	0.076	-0.326	3.456	0.552
RYWH	0.069	-0.180	2.765	0.841
RYCGX	0.061	-0.266	2.932	0.763
RYWHX	0.114	-0.769	3.486	0.087
RYCGEU	0.051	-0.532	3.812	0.187
RYWHEU	0.064	0.964	5.285	0.000
RYOLSEU	0.093	0.063	2.624	0.921
RYSBAB	0.092	0.101	2.710	0.901
RYCGOXE	0.035	-0.059	2.931	0.983
RYWHOXE	0.063	0.108	3.639	0.653
RYOLSO	0.076	0.178	2.136	0.721

[a] For RYCGOCN and RYSBAB, the years are 1965 to 2004.
For RYOLSEU and RYOLSO, the years are 1980 to 2004.

Fig.1 Histogram and Cumulative Distribution of US Corn Yields in 2006 from 1000 Random Draws from Actual Deviations in 1960 to 2004

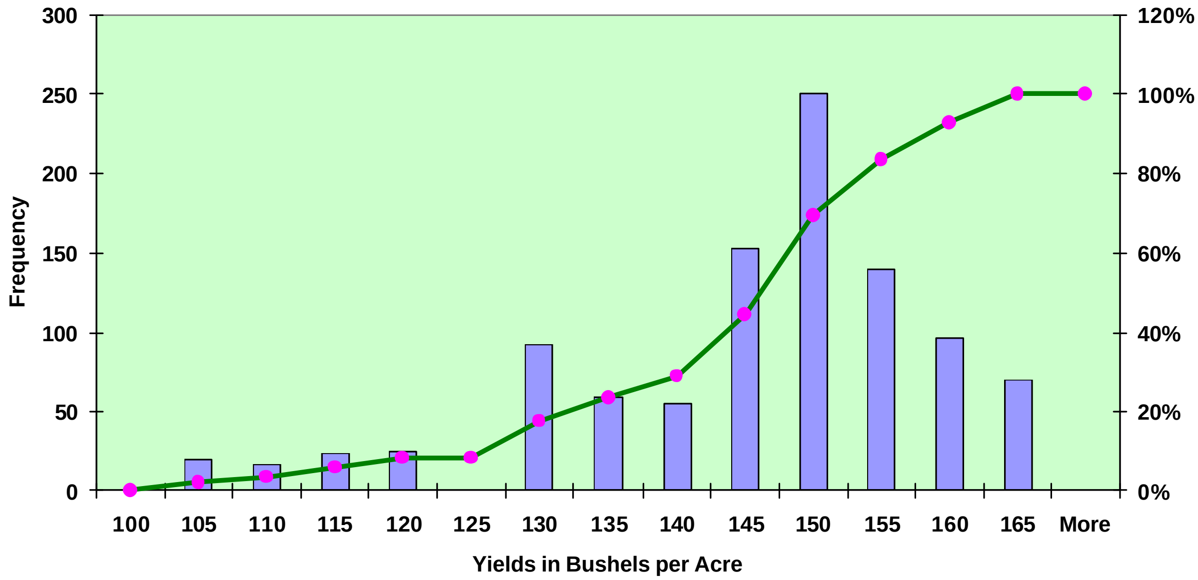

Fig.2 Histogram and Cumulative Distribution of US Soybean Yields in 2006 as a Function of US Corn Yield Deviations and a Normal Distribution

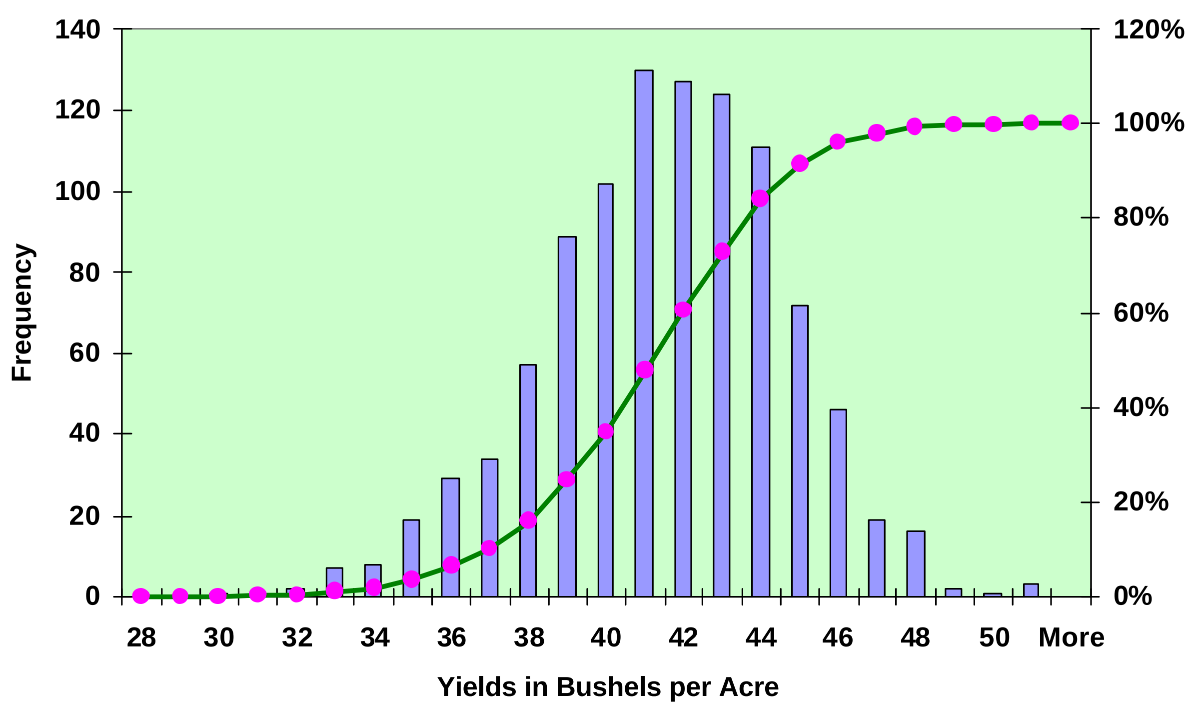

Fig.3 Histogram and Cumulative Distribution of US Wheat Yields in 2006 based on 1000 Draws from a Normal Distribution with Wheat's Historic S.D.