

# **A Mixed Effects Model of Crop Yields for Purposes of Premium Determination**

**Michael J. Roberts<sup>1</sup> and Jesse B. Tack<sup>2</sup>**

*Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2011*

*AAEA & NAREA Joint Annual Meeting, Pittsburgh, Pennsylvania, July 24-26, 2011*

*Copyright 2011 by Michael Roberts and Jesse Tack. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.*

---

<sup>1</sup> Department of Agricultural & Resource Economics, North Carolina State University, michael\_roberts@ncsu.edu

<sup>2</sup> Department of Agricultural Economics, Mississippi State University, tack@agecon.msstate.edu

## Overview

Farm income is highly variable due to annual price and yield uncertainties. The federally subsidized crop insurance program is an important tool for managing this risk, and has grown from a relatively modest program to one that encompasses the majority of productive cropland in the country. In most years since 1995, farmers have enrolled over 200 million acres - about two thirds of all U.S. cropland. In nominal dollars, total liability insured under the program steadily increased from almost \$6 billion in 1981, to nearly \$24 billion in 1995, to over \$40 billion in 2003, to almost \$90 billion by 2008. Indemnities have also increased, from about \$1.2 billion in 1989 to almost \$2.4 billion in 1999 to over \$8.6 billion by 2008.

The success of this program depends on identification of actuarially fair insurance premium rates, which in turn depends on accurate estimation of farm-level yield distributions. Previous research has used a wide range of parametric, semi-parametric and nonparametric approaches to model and estimate yield distributions. A common limitation of these studies is the use of county-level NASS yield data and/or not fully accounting for how location, time, and farm management might shift the distribution.

We use the confidential U.S. Department of Agriculture Risk Management Agency (RMA) panel dataset to estimate farm-specific distributions of yields and actually fair crop insurance premiums. We consider coverage in eleven states that grow a significant portion of the nation's five largest crops (in terms of production value): corn, soybeans, wheat, cotton, and rice. This dataset consists of millions of field-level yield observations of all Federal crop insurance participants from 1992 to 2002, which presents a unique opportunity to characterize the sources of heterogeneity for farm-level yield distributions. The wealth of data allow for precise estimates of the entire distribution of yield outcomes even while using a large number of fixed and random

effects to control for unobserved factors relating to farmer ability, land quality, technology, weather, and local prices.<sup>3</sup> Since this approach uses farm-level data and incorporates several important sources of heterogeneity, it likely provides more precise estimates than previous research.

Our ongoing work includes using the difference between our estimated actually fair premiums and RMA's to predict which insurance contracts farmers select. Ultimately, we will predict potential efficiency gains from using our empirical model for premium determination.

## Research Methodology

The goals of this research include empirical estimation of farm-level yield distributions, calculation of actuarially fair risk premiums, and prediction of potential efficiency gains from using our empirical model for premium determination. We assign a distribution to each farm-specific yield given the available information received by the RMA. We expect farmers may differ not just in their expected yield, but also their variance. We also expect farmers in the same county have similar yield distributions and may experience similar shocks each year, but also have an idiosyncratic component unique to each farmer.

Consider the model:

$$y_{it} = a_i + b_c t + c_c ins + u_{ct} + \varepsilon_{it} \quad (1)$$

where  $y_{it}$  represents yield on farm  $i$  in year  $t$ ,  $a_i$  is a farm-specific intercept,  $b_c$  is a county-specific trend,  $c_c$  is a county-specific insurance effect,  $u_{ct}$  is a county-specific random shock in year  $t$ , and  $\varepsilon_{it}$  is an idiosyncratic random shock on farm  $i$  in year  $t$ . The farm-specific intercept

---

<sup>3</sup> The inclusion of both fixed and random effects in an econometric model is referred to as a mixed effects model. This framework is gaining popularity for empirical economic analyses that use data with a clustered structure

can be written as a county-specific mean plus a farm-specific shock,  $a_i = a_c + u_i$ , and the county-specific slope parameters can be written as a population mean plus a county shock,  $b_c = b + u_{c1}$  and  $c_c = c + u_{c2}$ . These modifications generate the model:

$$y_{it} = a_c + bt + cins + u_i + u_{ct} + u_{c1}t + u_{c2}ins + \varepsilon_{it} \quad (2)$$

There are two types of effects in this model. The first part,  $a_c + bt + cins$ , represents the fixed effects of crop yields, and the second part,  $u_i + u_{ct} + u_{c1}t + u_{c2}ins + \varepsilon_{it}$ , represents random effects. There are two random intercepts, farm-specific ( $u_i$ ) and county/time-specific ( $u_{ct}$ ); and, there are two random slopes at the county level, for the trend parameter ( $u_{c1}$ ) and for the insurance parameter ( $u_{c2}$ ). We follow conventional methods for estimating linear mixed models by assuming that the random components follow a multivariate normal distribution and employ maximum likelihood. Although not discussed here, we use a second stage to estimate a mixed effects model for the variance of the idiosyncratic term  $\varepsilon_{it}$ , which allows us to identify fixed and random sources of heteroskedasticity for the yield model.

We estimate a separate model for each state-crop combination in our dataset, and empirical results suggest that there are several important sources of heterogeneity for crop yield distributions. For example, for Arkansas rice, the farm-specific random intercept accounts for 42 percent of the random variation of the model, the county-specific intercept accounts for 7.5 percent, the random slope for the time trend accounts for 0.1 percent, and the random slope accounts for 1 percent. These estimates confirm our suspicion that there are significant farm-level crop distribution shifters across space and time.

Ongoing work includes using the estimated model to evaluate the efficiency of RMA's crop insurance program. This research is likely to generate several interesting findings: (i) the role of

linear mixed models in empirical analysis, (ii) particular sources of heterogeneity in U.S. crop yield distributions, and (iii) potential increases in efficiency for the federal crop insurance program.