Appendix II
The Logistic Regression Model

The report uses a logistic regression model to estimate the probability of farm exit (*P*) during each intercensus period, as in the following (Greene, 1993, p. 297):

$$\ln [P_{it}/(1-P_{it})] = Y = \beta' X_{it} + \varepsilon_{it} \qquad (1)$$

where ln is the natural logarithm, X is a vector of exogenous variables, (for example, various farm and operator characteristics) for the i*th* farm in time period *t*, β is a vector of parameters to be estimated, and ε_{it} is a stochastic error term. Coefficients in logistic regressions (the β parameters) tell how much a change in an independent variable changes the log of the predicted odds ratio $[P_{it}/(1-P_{it})]$. Because we are interested in the effects on the predicted probability of exit (P_{it}), we must derive the predicted probability as:

$$P = e^{Y}/(1+e^{Y}) = e^{\beta' X}/(1+e^{\beta' X}), \qquad (2)$$

where e is the base of natural logarithms, approximately equal to 2.718.

Linear regression models are inappropriate for our data because they may give nonsensical predicted probabilities for exit—exceeding 100 percent or less than zero. Logit or probit models are usually chosen for estimation in cases where the object is to analyze the choice between two alternatives, in this case, exit or continued operation. In cases like this one, where the explanatory variables are themselves dichotomous, the logit is likely to be preferred because the probit's assumption of normally distributed error terms may not be appropriate (Kennedy, 2003, p. 267).

Equation 2 indicates that the effect of changes in an explanatory variable on the probability of exit will be nonlinear and will vary with the values of other explanatory variables. For that reason, the report presents predicted exit probabilities, in tables 3, 4, and 6-9 and in figures 5-8, for different combinations of explanatory variables. To derive the predicted exit probabilities, we first estimated the logistic regressions to obtain the parameter estimates β. We then combined the parameter estimates with various representative values of the explanatory variables *X* to derive predicted values for Y, the log of the odds ration. For any given value of Y, the predicted exit probability *P* can be derived as $e^{Y}/(1+e^{Y})$.

The Base Model

Operator age and farm size are two fundamental determinants of exit. We first explored a base model that uses only those determinants. We used this base model for three reasons. First, knowledge of exit probabilities across various size and age categories is useful in itself. Second, the exact linkage between age and size to exit may be complex. Because we wanted to explore potential nonlinearities using categorical measures, we did not want to complicate the model more by adding additional variables. Third, the base model estimates provide a useful point of comparison when we add additional variables.

In the base model, we used dummy variables that depict four age classes and six size classes, with size measured in sales, adjusted for inflation with the Producer Price Index for Farm Products:

Age classes	***Sales classes***
Years	*1997 dollars*
Younger than 45	Less than 1,000
45-54	1,000-9,999
55-64	10,000-49,999
65 or older	50,000-99,999
	100,000-249,999
	250,000 or more

Tests of the Base Model Specification

The base model was selected from three potential logit models that were evaluated for significance in predicting a farm's exit. The two rejected alternatives were as follows:

- **Sales cubed, age squared.** We replaced the categorical sales categories with continuous measures, using sales, sales squared, and sales cubed as well as age and age squared (we also used continuous sales measures with age classes). This alternative provided a weaker fit to the data, however, compared with using sales and age classes.
- **Four age classes, six sales classes, and their interaction terms.** Including interaction terms in the third model helps determine whether there are combination effects among the variables. This combination results in a less significant log likelihood than the second model and produces several insignificant t-statistics. We found no evidence of improved fit from adding the interaction terms.

All the models tested produce highly significant t-statistics. Highly significant t-statistics are to be expected because the longitudinal data base is so large (4.5 million observations). The huge underlying data set used in this report—coupled with the long time span between census years (generally 5 years)—also should help alleviate the effects of possible econometric problems.

Additional Models

Once we accepted a base model, we constructed five other models by adding measures of race, gender, specialization, off-farm work, and business age. Our goal was to use the logit model to estimate exit probabilities, controlling for size and age. We felt that developing this approach was important because size and operator age varies sharply across the categories in the other explanatory variables.

The coefficients from each logistic regression are presented in appendix table 2. No coefficients are presented for the 1978-82 intercensus period. Unlike the other periods, it is only 4 years long (rather than 5), and coverage of very small farms is incomplete in the 1978 Census. (See appendix III for more information about coverage in the 1978 Census.)

Appendix table 2

Logistic regression coefficients by intercensus period[1]

Model and variables		1982-87	1987-92	1992-97
Base model				
	Intercept	-0.533	-0.750	-0.652
Sales:	Less than $1,000	.903	1.114	.916
	$1,000-$9,999	.515	.686	.553
	$10,000-$49,999	.243	.424	.342
	$50,000-$99,999	.148	.229	.261
	$100,000-$249,999	-.046	.047	.153
Operator age:	Younger than 45	-.220	-.290	-.398
	45-54	-.435	-.431	-.522
	55-64	-.314	-.307	-.395
Value of log likelihood function		*-1,336,652*	*-1,185,219*	*-1,064,789*
Race model (excluded category: White)				
	Intercept	-0.542	-0.759	-0.658
Sales:	Less than $1,000	.894	1.111	.914
	$1,000-$9,999	.509	.685	.552
	$10,000-$49,999	.243	.427	.343
	$50,000-$99,999	.150	.232	.263
	$100,000-$249,999	-.044	.051	.155
Operator age:	Younger than 45	-.214	-.287	-.396
	45-54	-.431	-.428	-.521
	55-64	-.312	-.305	-.394
Race:	Black	.409	.358	.266
	Native American	.187	.197	.095
	Asian	.314	.494	.383
	Other	.215	.298	.147
Value of log likelihood function		*-1,335,948*	*-1,184,684*	*-1,064,553*
Business age model (excluded category: 14 years or more)				
	Intercept	NA[2]	NA[2]	-0.809
Sales:	Less than $1,000	NA[2]	NA[2]	.771
	$1,000-$9,999	NA[2]	NA[2]	.477
	$10,000-$49,999	NA[2]	NA[2]	.323
	$50,000-$99,999	NA[2]	NA[2]	.278
	$100,000-$249,999	NA[2]	NA[2]	.187
Operator age:	Younger than 45	NA[2]	NA[2]	-.635
	45-54	NA[2]	NA[2]	-.646
	55-64	NA[2]	NA[2]	-.461
Business age:	Less than 5 years	NA[2]	NA[2]	.691
	5-9 years	NA[2]	NA[2]	.357
	10-13 years	NA[2]	NA[2]	.143
Value of log likelihood function		*NA[2]*	*NA[2]*	*-1,050,754*

Model and variables		1982-87	1987-92	1992-97
Specialization model (Excluded category: Other livestock)				
	Intercept	-0.666	-0.816	-0.749
Sales:	Less than $1,000	1.031	1.181	1.002
	$1,000-$9,999	.607	.808	.651
	$10,000-$49,999	.272	.496	.399
	$50,000-$99,999	.161	.268	.286
	$100,000-$249,999	-.037	.074	.167
Operator age:	Younger than 45	-.255	-.338	-.438
	45-54	-.454	-.462	-.551
	55-64	-.327	-.325	-.411
Type:	Cash grains	.214	.124	.126
	Other field crops	.222	.170	.205
	Vegetables and melons	.477	.340	.387
	Fruits and tree nuts	.157	.147	.148
	Horticultural	.708	.576	.538
	General crops	-.035	-.064	-.096
	Beef cattle	-.103	-.215	-.137
	Hogs	.152	.085	.278
	Dairy	.082	.002*	.084
	Poultry and eggs	.239	.263	.136
	Animal specialties	.200	.292	.173
Value of log likelihood function		*-1,330,749*	*-1,178,111*	*-1,059,921*
Off-farm work model (Excluded category: 200+ days)				
	Intercept	—[3]	-0.700	-0.612
Sales:	Less than $1,000	—[3]	1.096	.904
	$1,000-$9,999	—[3]	.671	.542
	$10,000-$49,999	—[3]	.417	.337
	$50,000-$99,999	—[3]	.230	.262
	$100,000-$249,999	—[3]	.049	.155
Operator age:	Younger than 45	—[3]	-.301	-.403
	45-54	—[3]	-.443	-.529
	55-64	—[3]	-.313	-.398
Days:	No days of off-farm work	—[3]	-.042	-.034
	1-199 days of off-farm work	—[3]	-.087	-.093
Value of log likelihood function		*—[3]*	*-1,185,021*	*-1,064,602*
Gender model (Excluded category: Male)				
	Intercept	-0.557	-0.774	-0.676
Sales:	Less than $1,000	.879	1.088	.885
	$1,000-$9,999	.495	.669	.533
	$10,000-$49,999	.235	.417	.333
	$50,000-$99,999	.146	.228	.259
	$100,000-$249,999	-.047	.047	.153
Operator age:	Younger than 45	-.200	-.274	-.383
	45-54	-.417	-.415	-.510
	55-64	-.301	-.295	-.384
Gender:	Female	.404	.376	.353
Value of log likelihood function		*-1,334,639*	*-1,183,421*	*-1,063,123*

Note: All coefficients are significant at the 99-percent level, except dairy in 1987-92.

* = Not significant.

[1]Excluded categories for all models: sales—$250,000 or more; operator age—65 or older.

[2]NA = Not applicable. The analysis was performed for only the 1992-97 period. It examines exits between 1992 and 1997, by business age.

[3]— = Not available. The longitudinal file has days of off-farm work from 1987 forward.

Source: 1997 Census of Agriculture Longitudinal File.