Modeling the Credit Risk in Agricultural Mortgages: A Critical Review of the Farm Credit Administration's Credit Risk Model for Farmer Mac

Austin Kelly
Assistant Director, Center for Economics
United States Government Accountability Office
441 G St. NW
Washington, DC 20548
(202) 512-7186
kellya@gao.gov

September 2004. Draft.

Note - The opinions expressed in this paper are solely those of the author and do not reflect the position of the United States Government Accountability Office.

Abstract

Farmer Mac is the GSE charged with creating a secondary market in loans backed by agricultural real estate. The Farm Credit Administration (FCA) has estimated a credit risk model for agricultural mortgages. This model is a key determinant of Farmer Mac's risk based capital (RBC) requirement. This paper reviews both the structure of FCA's credit risk model, and the data used by FCA's contractors to estimate the model. Serious concerns are raised about both data quality and the econometric specification in use. Under Basel II, RBC models will proliferate. Assessing the validity of credit risk models will become increasingly important.

Overview

Farmer Mac is a Government Sponsored Enterprise (GSE) charged with providing a secondary mortgage market for agricultural loans.[1] It is essentially the Fannie Mae or Freddie Mac for the farming sector, and enjoys similar GSE status and a similar funding advantage. Farmer Mac opened its doors in January 1988, created by the Agricultural Credit Act of 1987 (PL 101 - 233). Its powers were significantly expanded in 1996, and the bulk of its $5.5 Billion portfolio of loans and guarantees consists of post 1996 business. Farmer Mac retains in portfolio over 90% of the whole loans that it purchases. A growing segment of its business consists of Long-Term Standby Purchase Commitments, an off-balance-sheet item giving lenders the right to put delinquent loans with Farmer Mac for a payment equal to their unpaid principle balance.

The Farm Credit Administration is charged with regulating Farm Credit System institutions, including Farmer Mac. Thus FCA has responsibilities similar to OFHEO's, which it exercises through its Office of Secondary Market Oversight (OSMO). In fact, the law that mandates the risk-based capital regulatory regime for Farmer Mac (1991 Amendments to the Farm Credit Act of 1971, PL 102-237) generally parallels the language in FHEFSSA, the statute that governs OFHEO's regulatory regime. It specifies minimum leverage for on and off balance sheet items,[2] and a Risk-Based Capital (RBC) regime, measured as the sum of credit, market, and operations risk. The statute specifies that credit risk be measured in a worst case scenario, echoing the language in FHEFSSA, that market risk be measured by parallel shocks to the yield curve, that operations risk be a 30 percent add on, and that Farmer Mac be required to hold enough capital to withstand 10 years of a stressed environment, defined as the highest defaults over a 2 year period in contiguous areas which contain at least 5% of the US population. Unlike OFHEO, FCA was given the right (but not obligation) to consider new business profits (or losses) immediately as part of the risk-based-capital regime.

The remainder of this paper is as follows. First there is a brief discussion of the identification of the stress case and the identification of relevant data. This is followed by an overview of the modeling of credit risk, including the sources of data and structure of the credit risk model. The next sections discuss, in order, the logistic regression used to model foreclosure probabilities, the Beta distribution used to predict the timing of foreclosures, and the loss given default (severity) estimate in use. The ability of the model to predict credit losses out-of-sample is then discussed. The paper ends with some observations on the potential for improvements in the model.

Early Work

In response to the risk-based-capital statute, FCA contracted out a series of studies to identify the 2 year worst case credit loss scenario called for in the legislation, and to build a credit risk model for Farmer Mac's portfolio. FCA's consultants identified the Upper Midwest, Minnesota, Iowa, and Illinois, over the 1983-1984 period, as the scenario representing the worst case credit risk for agricultural mortgages. This was based on aggregate default information. (Barry, et al 1993.).

FCA's consultants also engaged in a reconnaissance of loan level data sources. Serious data availability problems were identified. Many Farm Credit System institutions used a centralized servicing facility, which shut down as a result of substantial losses during the mid 1980's farm crisis. This event resulted in the loss of historical loan level data for many FCS institutions. FCA's consultants identified two Farm Credit Banks, the Farm Credit Bank of St. Paul, and the Farm Credit Bank of Texas, as still in possession of a long history of loan level data. Although the St. Paul Bank was active in the highest stress region, the consultants felt that the Bank's reliance on restructuring as opposed to foreclosure, coupled with the complexities induced by temporary foreclosure moratoria imposed by states in the St. Paul region, limited the usefulness of St. Paul Bank data. Texas was the 3rd or 4th most stressed region identified, and its peak stressful period occurred 2 years later, in 1985-1986.

Hence, the strategy followed by FCA and its consultants was to use FCB Texas loan level data to build an econometric model of credit risk, and then to extrapolate the credit risk on a loan to the 1983-1984 Upper Midwest stress scenario. FCA's model is documented in the Code of Federal Regulations, 12 CFR 650, Appendix A.

FCA's Credit Risk Model

Data

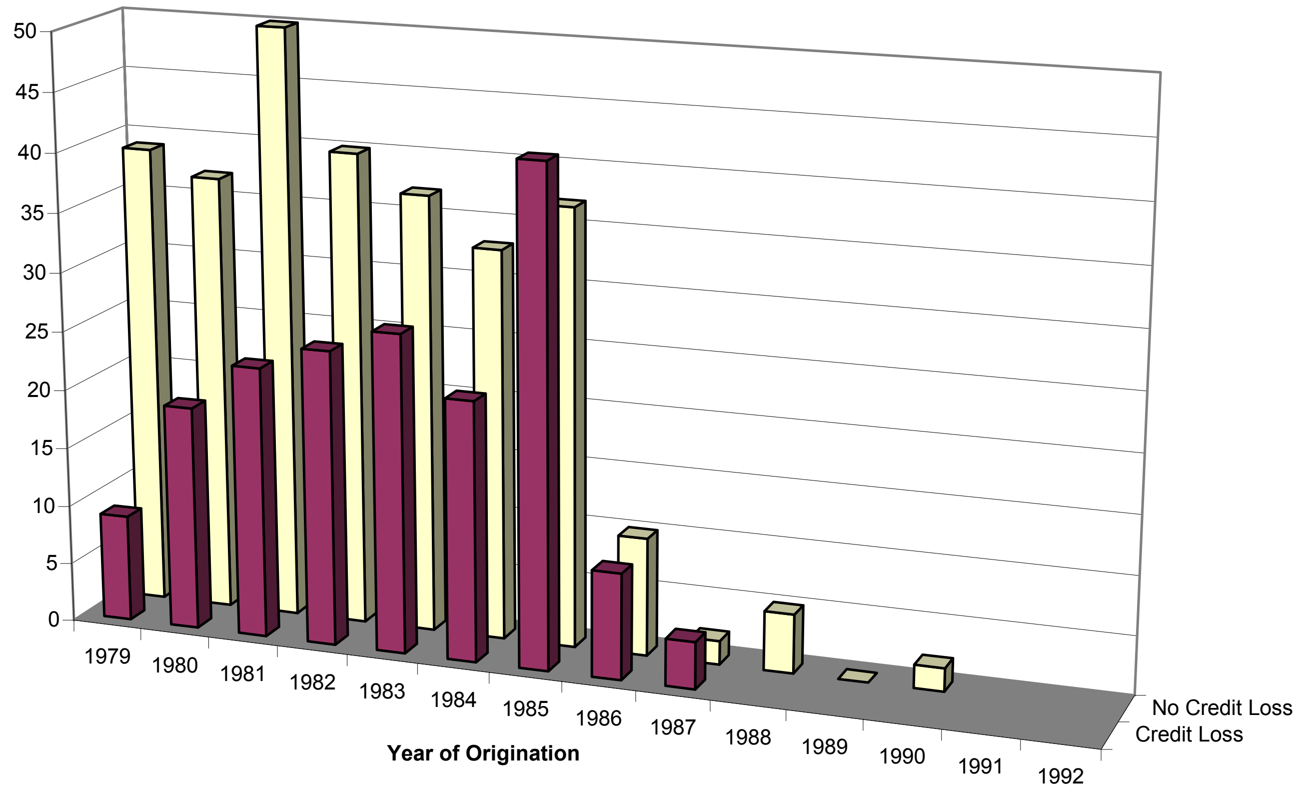
FCA's credit risk model is based on a file of 19,418 agricultural mortgage loans held by the Farm Credit Bank of Texas between 1979 and 1992 (see Fig. 1). This subset of data chosen for analysis consisted of loans which would have met, or fallen just outside of, Farmer Mac's purchase guidelines. The file included data on underwriting variables, such as loan size, loan to value ratio, debt to asset ratios, and a foreclosure indicator and foreclosure date, current as of the end of 1992. The file also included an estimate of net losses on foreclosed loans made at the time a loan entered the REO portfolio. There were 180 foreclosed loans which resulted in an estimated credit loss in the 19,418 loan file. All 180 of the loans which produced a credit loss foreclosed in 1986 or later, and only 4 of the loans originated post 1986 suffered a credit loss. These 4 loans were 1987 originations. The unconditional probability of a credit loss on loans originated prior to 1987 was about 1.5 percent, on 1987 originations about 0.5 percent, and was zero on post 1987 originations (see Fig. 2). A further 208 loans went through the foreclosure process, but no credit loss was recorded in the database.

Structure

The credit risk model consists of 3 equations. One predicts the probability of a loan going to foreclosure and resulting in a credit loss, one predicts the timing of the foreclosure, conditional on the loan ever foreclosing, and a third estimates the net credit loss on a foreclosed loan.

Fig. 1 Number of Loans by Origination Year

Fig. 2 Foreclosures by Origination Year



Foreclosure Regression

The foreclosure model is a logistic regression, with a dependent variable that identifies whether or not the loan had resulted in a credit loss by the end of the observation window in 1992. The independent variables consist of non-linear transformations of the initial loan-to-value ratio, the initial loan balance in 1997 dollars, and a land price "shock" variable, along with the ratio of the stock of debt to the stock of assets, and the debt service coverage ratio.

$$\text{Eq 1) Credit Loss} = \text{logit}\left(\alpha + ß1\ \text{LTV}^{\Gamma 1} + ß2\ (1\text{-exp}^{(\Gamma 2\ \text{LS})}) + ß3\ \text{PS}\ (1/(1+\Gamma 3))^{\text{Time}} + ß4\ \text{DA} + ß5\ \text{DSCR}\right)$$

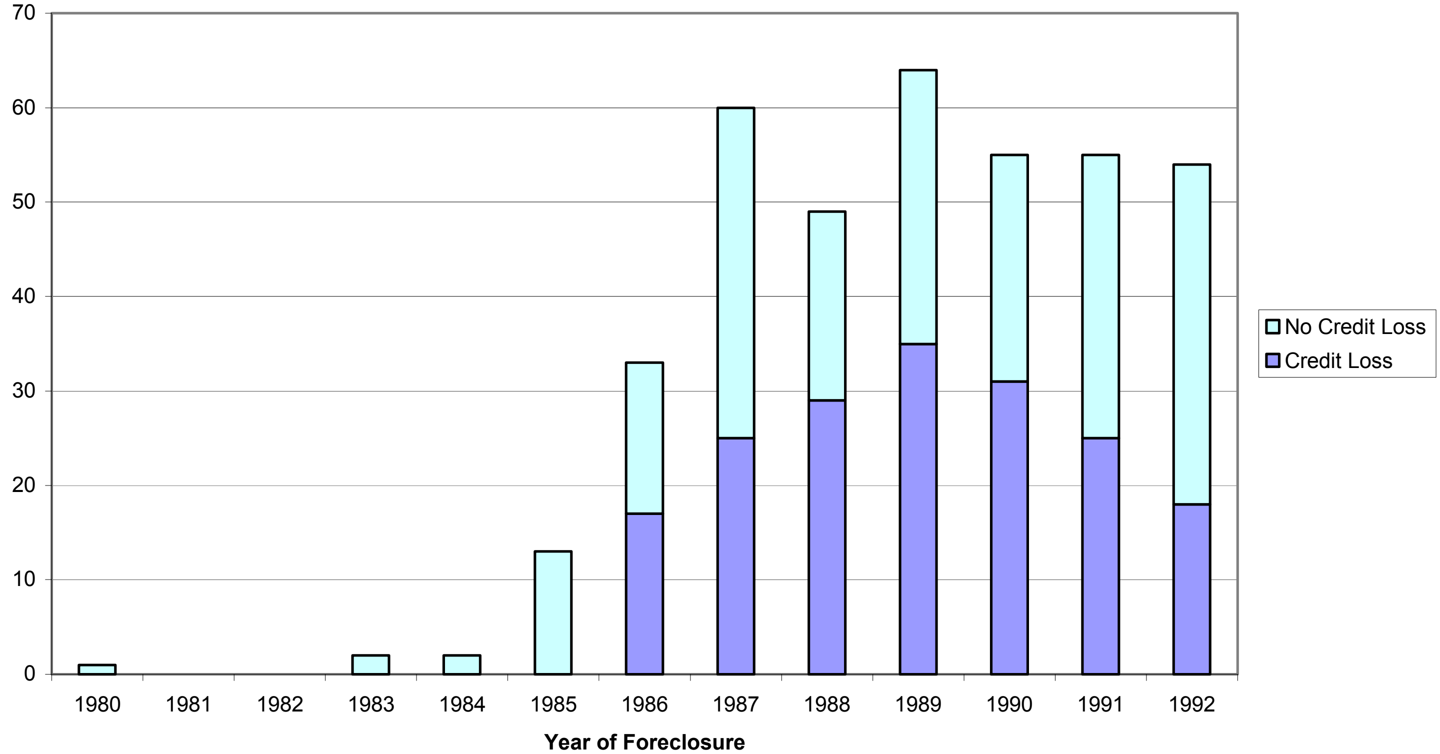
Where LTV is the Loan-to-Value ratio, LS is the Loan Size (in thousands of 1997 dollars), PS is the Price Shock, DA is the Debt-to-Assets Ratio, and DSCR is the Debt Service Coverage Ratio.

Table 1 – Predictor Variables

Variable	**Coefficient**	**Nonlinear transformation**	**Transformation Coefficient**
Intercept	α = -12.62738		
LTV	ß1 = 1.91259	Power	Γ1 = 5.3914596
Loan Size (thousands)	ß2 = 4.55390	Negative Exponential	Γ2 = -0.00538178
Price Shock	ß3 = - 0.33830	Dampening %	Γ3 = 0.0413299
Debt/Assets	ß4 = 2.49482	None	
DSCR	ß5 = - 0.19596	None	

There are several problems with both the structure of this equation, and the variables used as explanatory variables. The first structural problem is that there is no adjustment made for right censoring. It is likely that some of the loans that survived through 1992 have gone to foreclosure after 1992. While it is impossible to conclusively determine that this is the case without access to post 1992 data, it seems likely based on two facts. First, 18 of the total of 180 loans that

Fig. 3 Foreclosures by Foreclosure Year

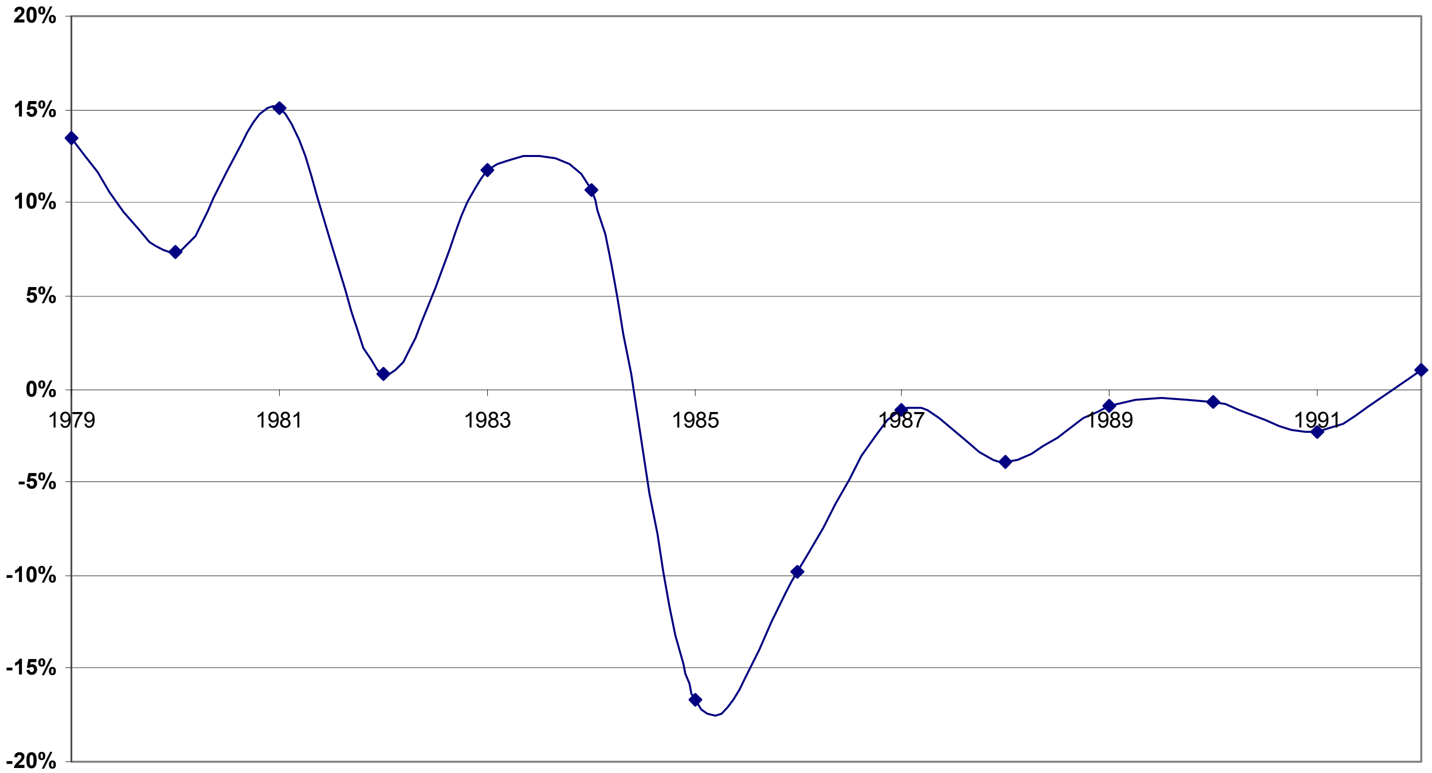


foreclosed with a credit loss completed the foreclosure process in 1992 (see Fig. 3), and origination years on these loans ranged from 1980 to 1987. Second, the peak foreclosure year for these loans is the 6^{th} and over one-quarter of the loans in this population have not yet reached their 6^{th} year. Thus there is room for a substantial number of loans to have resulted in a credit loss that remains unrecorded in this dataset. Most models of credit risk use a hazard rate framework to estimate annual (or quarterly, etc.) loan termination probabilities, and to project those rates forward for loans in mid-stream, in order to solve the right censoring problem.[3] The use of dummy variables to identify origination years can also be used in the absence of time-varying variables. Failing to account for right censoring would generally lead to an underestimate of lifetime foreclosure probabilities.

The unique characteristics of this dataset may lead to a further problem caused by the failure to deal with right-censoring. Older loans are much less likely to suffer from an undercount of foreclosures. Because the land price "shock" in this dataset occurs in the very middle of the observation window, age of loan is strongly correlated with the land price shock variable, and with the measurement error in the dependent variable (see Fig. 4). Thus, right-censoring may lead to a biased coefficient as well as an undercount (at mean values of the independent variables) of foreclosure propensities.

The use of an equation to predict unconditional foreclosure rates, rather than conditional foreclosure rates (hazard rates), leads to a further problem. The unconditional lifetime foreclosure rate on a cohort of loans may strongly depend on the prevalence of the competing risk of loan termination, prepayment (Foster and VanOrder 1985). Consider, for example, a cohort of loans originated in 1982, near the peak of interest rates. Falling rates between 1982 and 1985 induced a large fraction of this cohort to prepay their mortgages prior to the end of 1985. Thus, much of this cohort would not have survived to experience the 1985-1986 land price shock in Texas. A regression estimated on loans that had passed through a refinancing wave, but applied to a new cohort of loans that will not experience a strong refinancing incentive, may seriously underestimate the number of defaulting loans, because more loans in the new cohort will survive long enough to experience a shock event. Additionally, the regression is estimated on all loans originated, but the default probability is calculated only for all loans that

Fig. 4 Texas Farmland Price Changes



have survived to the date at which the capital test is run - a left censoring problem for the application of the risk based capital test to a portfolio of seasoned loans.[4]

Again the unique characteristics of this dataset may serve to amplify the problem that arises from ignoring the competing risk. The typical Farm Credit System loan written in the 1970's and early 1980's was an adjustable interest rate loan, with the interest rate tied to the Farm Credit System's cost of funds, similar to the FHLB's Cost Of Funds Index (COFI). Post 1979, the Farm Credit System was able to capture a large share of the agricultural market, because its cost of funds index was, essentially, a weighted average of past and current interest rates. At a time of sharply rising rates, FCS institutions offered lower interest rates than could competitors, who presumably priced off the marginal, not average, cost of funds. However, post 1982 this situation reversed itself. As interest rates fell, FCS's cost of funds index continued to rise, as old debt issued at very low rates continued to roll off, being replaced with new debt at higher rates. Thus, the adjustable rates on FCS loans continued to rise post 1982, while the rates offered by FCS competitors began to fall rapidly. The prepayment wave triggered by this chain of events led to severe financial stress at FCS institutions (US GAO 1985).

Post 1986, most FCS loans (and most of the loans in Farmer Mac's portfolio) are either fixed-rate loans with substantial prepayment penalties, or rapidly adjusting adjustable rate loans, tied, for instance, to 6 month LIBOR. Neither type of loan is likely to experience an interest rate induced wave of refinancings. Therefore, the reduction in credit risk which stems from prepayment (calculated over the entire life of the loan) is likely to be smaller for current books of business than it was for the loans used in the estimation sample.[5]

Variables

The dependent variable in the foreclosure regression is a 0-1 indicator for whether the loan generated a credit loss. There are 388 loans in the file that completed foreclosure by the end of 1992. However, only 180 loans were recorded as generating a loss, and it is this subset of 180 loans that are predicted by the regression. The loss variable used to identify this subset is an estimate of projected losses made at the time the loan entered REO. According to FCB Texas officials, it was an accounting estimate, and did not include opportunity costs, such as foregone interest during the REO process, which often took 1 to 2 years. It is also unlikely that Bank officials would have anticipated that prices for agricultural land would continue falling during the REO process (which they did for most years of the late 1980's). While it is possible for rational borrowers to default when the value of the collateral exceeds the value of the first lien, for example, if the combined effect of a first and a second is to more than exhaust the value of the collateral, the percentage of foreclosures that resulted in no credit loss, 62 percent, seems unusually large in comparison to multifamily mortgages, which are also commercial in nature.

Additionally, work outs and restructurings were excluded from the credit loss variable. While the Texas Bank used these tools much less often than did the St. Paul Bank, they were used on occasion. To the extent that loans were restructured on non-concessionary terms and the resolutions were successful, this may be appropriate. However, this treatment is not correct if the restructured loans later defaulted or if the loans were restructured on concessionary terms that represented a financial loss to the institution. The proper treatment for restructurings on concessionary terms depends on the answer to a taxonomic issue as to whether the losses should be labeled credit or market risk. From an accounting perspective, the unpaid principle balance is ultimately recovered, but the market value of a loan restructured on concessionary terms may be sharply reduced. Unfortunately, neither FCA's credit risk model nor its market risk model would capture this risk, as these models are currently structured.

The key independent variable in the foreclosure regression is the "land price shock" variable. It is defined as the smallest increase in Texas farmland prices observed between the year of origination and the year of foreclosure termination, or 1992, whichever comes first. It is designed to capture the impact of the -17 percent change in Texas farmland prices between 1985 and 1986, and is the variable upon which the extrapolation to the worst case (Upper Midwest shock of –23 percent) scenario mandated by legislation is based. Unlike many other models of loan performance, no other measure of post-origination price appreciation is used, so that a 3 year old loan in a state with 2 years of appreciation at 10%, followed by a 17% downshock, would be treated as having the same risk as a 3 year old loan with 2 years of 10% depreciation, followed by a 17% downshock.

This definition of the "land price shock" variable leads to several problems, both logical and structural, in the regression estimation. One problem stems from the USDA land price series used to measure land price changes,[6] and the timing incorporated in the construction of the observation matrix. The USDA series is an estimate of land prices for farmland made in the first quarter of the calendar year. The regression uses the change in, say, 1st quarter of 1985 to 1st quarter of 1986 to explain foreclosure for loans originated in 1986, although most or all of the shock occurred prior to origination. Because only foreclosure terminations and the right censoring date (1992) are used in the construction of this variable, it is also possible that the value assigned to a particular loan is for a shock that occurred after the loan terminated via prepayment. For example, a loan originated in 1980 and prepaying in 1984 would be assigned the 1986 land price shock value, because non-foreclosure termination dates are not used in the construction of this variable.

The definition used also leads to a serious structural problem for the modeling effort. The definition, greatest (in absolute value) decline in land prices from origination to foreclosure termination or 1992, uses the event modeled (foreclosure termination) in the definition of the "independent" variable. Yamaguchi (1991, pp.3-6 and pp.26-27) cautions against the use of explanatory variables in unconditional probability models if the value of the variable is not determined prior to an observation becoming at risk for the event being modeled. An important reason for using a conditional, as opposed to an unconditional, framework, is the former's ability, and the latter's inability, to incorporate time-varying covariates.

The definition of the land price shock in this particular dataset had some unusual ramifications for the estimated data matrix. Virtually[7] all loans originated before 1987 were assigned a value of –17% (the 1st quarter of 1985 to 1st quarter of 1986 value) while 1987 originations were assigned a value of -10% (the 1st quarter of 1986 to 1st quarter of 1987 value), and post 1987 originations were assigned small negative numbers (in the range of -2% to -4%) (see Fig. 5). Over 95% of the loans in the dataset had a land price shock value of either -17% or a small negative number. All loans that resulted in a credit loss had a land price shock value of either -17% (98% of credit losses) or -10% (2% of credit losses) (see Fig. 6). Essentially, the regression predicts the slope between about -3% and -17%, with all of the loans near -3% having a foreclosure rate of 0. There is almost no variation on the key predictor variable to use in ascertaining the

Fig. 5 Number of Loans by Price Shock

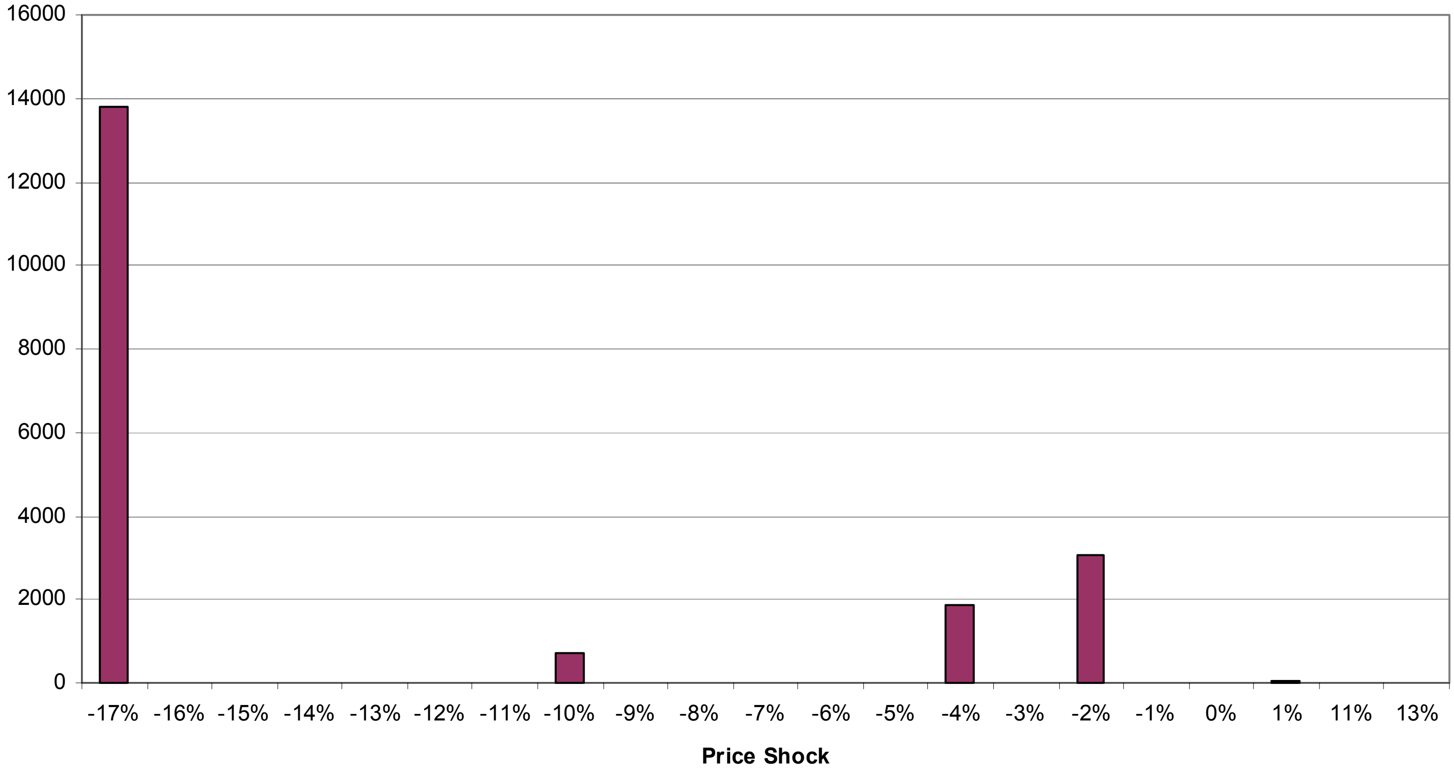
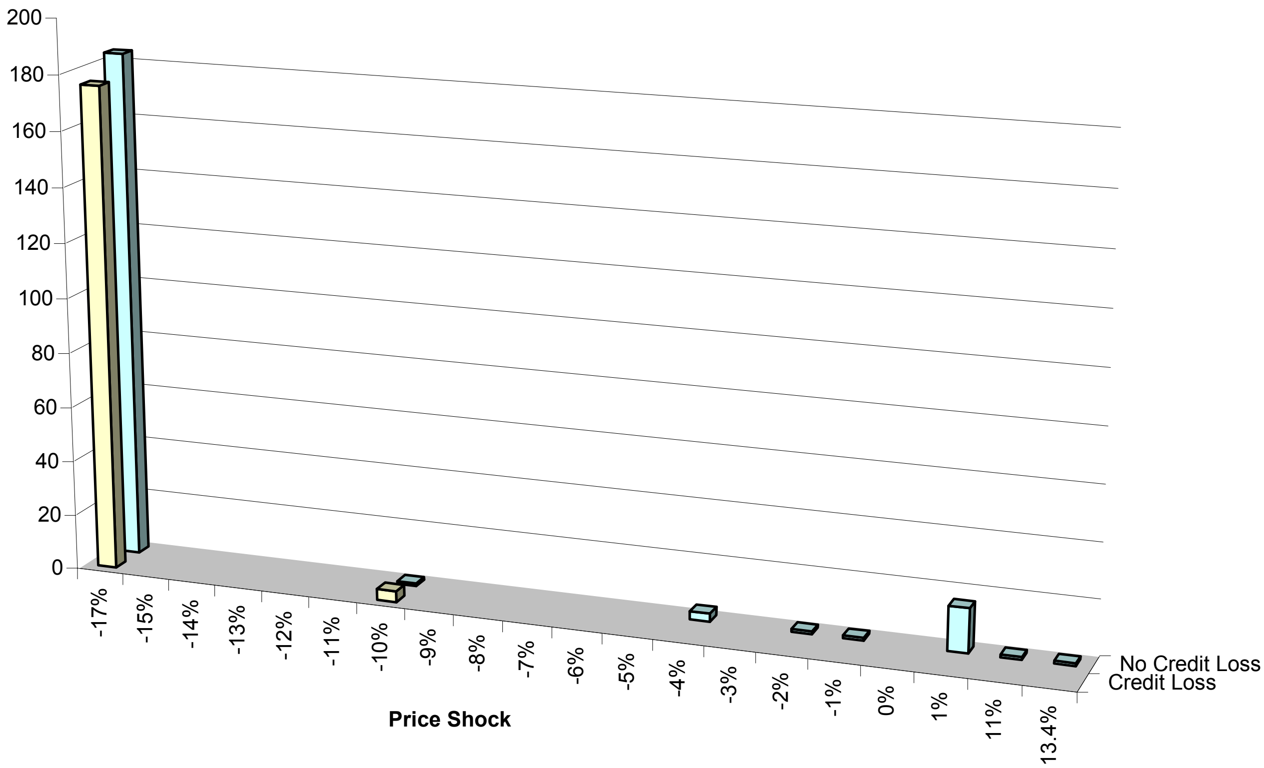


Fig. 6 Foreclosures by Price Shock



appropriateness of the assumed functional form. To examine the practical impact of the endogeneity of the key independent variable, the data used in the credit risk model was modified. The dependent variable was replaced with the result of a random number generator, the SAS Ranuni function, designed so that about the same number of credit losses (180) would be simulated by the random number generator as were in the original dataset. This was done once under the assumption that annual default probabilities were equal (so that a loan observed from 1979 to 1992 had 14 times the probability of a default as did a 1992 loan) and again with the probability of a loan defaulting in the first 2 years set to 0, the probability of the 13th and 14th year set to 0, and equal annual default probabilities for the remaining exposure years. The independent variable for the land price shock was then recalculated, and the regression re-estimated. A macro did this for 1000 times, for seed values fed to the random number generator of 1 to 1000. In both cases, 100% of the results showed a negative coefficient, significant at 5%. Coefficients on the other independent variables frequently changed sign and were significant less than 7% of the time.

Non-Linear Transformations

Three of the independent variables were transformed prior to their use in the logistic regression. Each variable was altered with a different functional form, and all functional forms and associated parameters were estimated outside of the logistic regression.

The land price shock variable was modified by a “dampening factor” which reduced the magnitude of the shock variable by 4 percent per year, with years measured from year of origination to year of land price shock or foreclosure. As is the land price shock variable itself, this dampening factor is endogenous: a 1979 loan that does not foreclose will be “dampened” by about 25% while a 1979 loan that forecloses in 1980 will be “dampened” by only 4%. The dampening factor is apparently designed to recognize the possibility that a seasoned loan may be less affected by a shock than a newly originated loan, perhaps because prepayments mean that the seasoned loan is less likely to have survived to the date of the shock, or because land price increases prior to the shock will tend to compensate for the shock. Since prepayments and land price changes have varied substantially over time, and are observable (at least at a high level of aggregation), a constant 4% “dampening” may be inadequate to capture these phenomena.

The LTV ratio at time of origination was entered as a power transformation, so that the variable in the regression is actually LTV raised to the power of 5.39. The power transformation has the effect of increasing the marginal effect of a change in LTV as the LTV gets larger. The loan balance at time of origination (in thousands of 1997 dollars) is entered as a negative exponential, so that the loan balance independent variable is actually 1 minus e raised to the (loan balance times –0.0054). This form produces a sharp bend in the estimated impact, with increases in loan size increasing risk at a diminishing rate up to about $500 thousand, and loan size increases causing almost no change in risk for values beyond $500 thousand. For example, the default probability increases by 50% as loan size increases from $100 thousand to $400 thousand, but by only 5% as loan size increases from $500 thousand to $10 million.

These non-linear transformations are estimated outside of the logistic regression model. Values for the non-linear transformation coefficients were first selected to produce the largest value of the likelihood function. The variables were then transformed and used as predictors in the logistic regression, and it is these coefficients and associated T-statistics that are reported in the Federal Register document. The estimation of these parameters outside of the logistic regression procedure renders the goodness of fit statistics reported with the model suspect (Kennedy 1987, p. 164). The endogeneity of the land price shock variable and associated dampening factor also violates the classical assumptions underlying the calculation of the goodness of fit statistics.

Omitted Variables

Additionally, the definitions of the two equity variables may incompletely represent the concepts of initial LTV and shock to equity. The FCB Texas loans used in the logistic regression were all Farm Credit System loans. FCS is structured as a cooperative. Borrowers using the system must purchase stock in their local lender, and this stock cannot be redeemed until the loan is paid-off. In effect, the stock purchase at time of origination can act as extra equity. Prior to 1986, stock purchase requirements were generally 10% of the initial loan balance. Stock purchase requirements were substantially eased in the mid 1980's, with typical requirements in the range of 2% to 5%. Therefore, most of the loans at risk of foreclosure would have had substantially more stock serving as an equity cushion than would the later loans in the dataset, or the FCS loans now purchased by Farmer Mac. And the non-FCS loans purchased by Farmer Mac would not have any stock purchase requirement.[8]

The land price shock variable does not encompass all the shocks to equity contemplated by the stress scenario. The stress scenario encompasses both a credit shock and stressful parallel shifts to the yield curve. Most fixed rate loans now purchased by Farmer Mac have substantial prepayment penalties, often taking the form of yield maintenance.[9] In many cases, their current pay-off amounts may substantially exceed their unpaid principle balances, as interest rates have generally been declining (even in the case of flat interest rates, the penalty may be over 5%). In the case of interest rate declines occurring at the same time as a land price shock, the fixed rate loans in the portfolio would experience a substantial increase in their pay-off amounts at the same time that the value of the collateral declined. To the extent that erosion of equity[10] from prepayment penalties led to foreclosures and recoveries less than the book value of the mortgage it would seem appropriate to treat this as an additional source of credit losses. To the extent that it might lead to the failure to collect contractually stipulated prepayment penalties it offers another point of debate over whether to classify these losses as credit losses or market losses,[11] but it is clear that this source of stress should be captured in either the credit risk or market risk models.

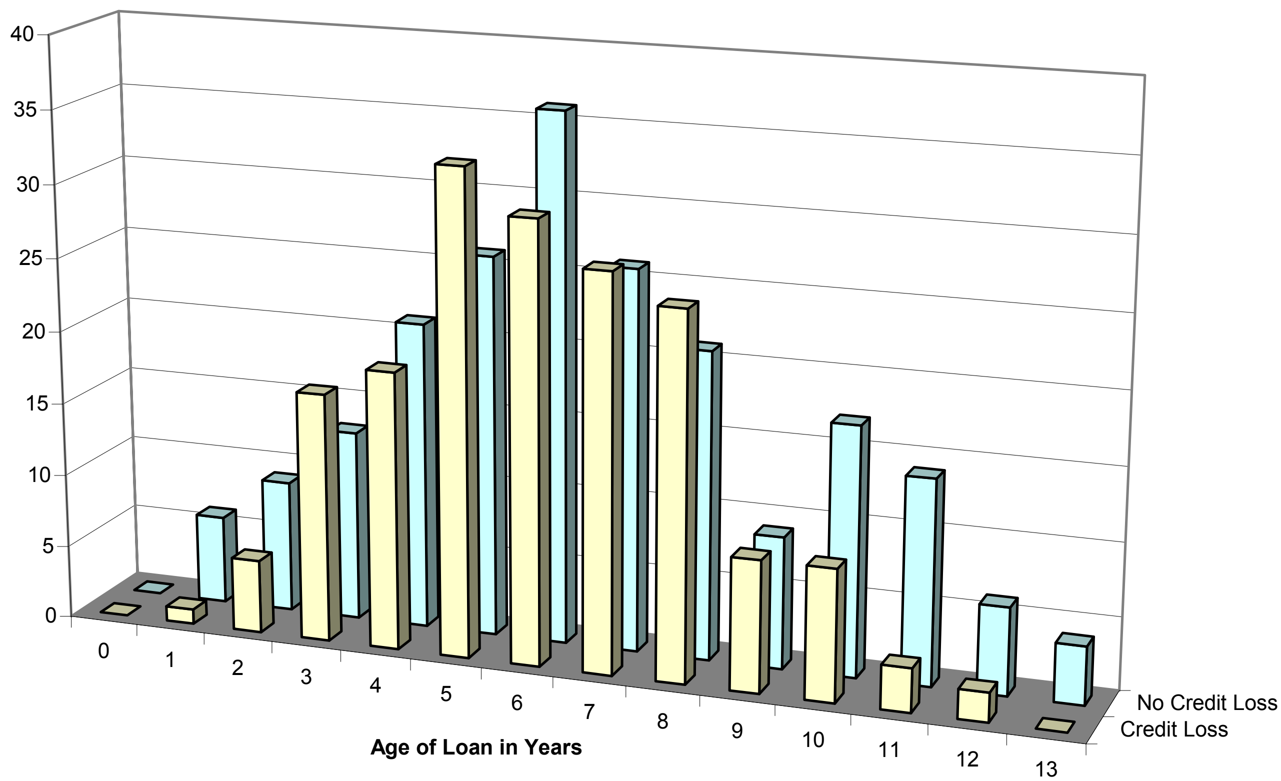
Beta Distribution

Because the credit risk regression produces a lifetime probability of default rather than an annual probability of default, some mechanism is required to distribute these projected losses over time. The technique used in FCA's model is to fit the two parameters of the Beta distribution to the distribution of observed foreclosure times in the FCB Texas data (see Fig. 7), assuming that the last year in which a loan can enter foreclosure is year 14.

Eq 2) Beta Dist $f(x/14) = \Gamma(\pi+\theta)/[\Gamma(\pi)\Gamma(\theta)] * x^{\pi-1}*(1-x)^{\theta-1}$

With $\pi = 4.288$ and $\theta = 5.3185$ and Γ representing the Gamma function

Fig. 7 Foreclosures by Duration



No correction is made for the right-censoring of the FCB Texas data at 1992. In general, this technique will lead to a bias towards an early mean for foreclosure dates, as foreclosures occurring early in loan lives will be over-represented. All foreclosures occurring in the first year will be captured in the data, but foreclosures occurring, for example, in the 10^{th} year, will only be recorded for loans originated in 1983 or earlier. A simple simulation was run, with origination years from 1979 to 1992, and with foreclosures following a triangular distribution between the 2^{nd} and 12^{th} years, with a peak foreclosure probability of 0.005 in the 7^{th} year. If the observation window runs at least 12 years past the last origination date, the simulation returns a mean time to foreclosure of 7 years. But if the last observation occurs 5 years after the last origination date, the mean time to foreclosure is only 6.6 years, and if the last observation occurs in the same year as the last origination date (as is the case with the dataset used in the FCA model) the mean observed time is only 6.4 years.

Additionally, the parameters of the Beta distribution are estimated independently of the risk parameters in the logistic regression. Most credit risk models take the form of a proportional hazards model, in which a baseline hazard rate (which serves a purpose analogous to the Beta distribution in the FCA model) is estimated jointly with the coefficients of the independent variables[12]. The fact that the Beta distribution is estimated independently of the risk factors renders its parameter estimates sensitive to the timing of shock events in the data. If a shock event occurs near the end of a data set all surviving loans will experience an increase in foreclosures, whereas if a shock occurs near the beginning of a data set loans only the earliest loans, with the longest observation windows, are affected. For example, a simulation similar to the above was run, and loans originated before a "shock" date had their annual foreclosure probability tripled after the shock date. If the shock date occurred late in the observation window, in this case 1990, there was almost no effect on the estimated times to foreclosure – the mean rose from 7 to 7.01. But if the shock occurred near the beginning of the observation window, at 1980, mean foreclosure time rose to 7.24. For a shock similar to that in the dataset used in the FCA model, occurring in 1986, mean foreclosure time rose from 7 to 7.11.

Severity

The final step in the credit risk model is the assignment of a severity rate to the loans that are predicted to default. The model uses the average severity observed in the loans in the Texas data that foreclosed with an estimated credit loss. That severity is 20.9%, which is roughly the weighted mean of the ratio of losses to loan balance at origination for all 180 loans which had a non-zero loss. Losses are not modeled as a function of underlying variables.

EQ 3) LGD = 0.209

As noted previously, the severity measure is based on an accounting estimate made when the loan entered REO. Important economic losses, such as foregone interest, were not captured. At the time when most losses were occurring, the late 1980's, interest rates on agricultural real estate were near 10 percent, and time in the REO inventory averaged between 1 and 2 years. Additionally, prices continued to decline for most of the post 1986 period, which may have lowered the prices actually realized for foreclosed property. The extent to which bank officials had incentives to accurately anticipate future losses during a financial crisis is unclear.

To the extent that severity may have been underestimated in the data, adjusting for these issues would tend to increase measured severity on the 180 loans with credit losses in the data. To the extent that severities are underestimated, and the foreclosures without credit loss in the model estimation are actually foreclosures with credit losses, the severity rate may be overestimated, as some loans with low severities may be excluded from the calculation. While loss given default may be either over or underestimated based on the above, the product of the default rate and the loss given default is clearly underestimated.

Fit

Fortunately, there has been no stressed time period for which stress results can be compared to an out of sample prediction from the model. However, the model's estimate of the slope of the price-shock -> credit loss relationship is anchored at two points, price shocks –3% and price-shocks of about –17%. Current experience does allow a test of the predictive power of the model in the unshocked range, by comparing model predictions with aggregate credit losses reported by Farmer Mac.

Since loan level data for Farmer Mac's portfolio are not public, assumptions must be made about the distribution of risk characteristics for the loans in their portfolio. For purposes of this comparison, a worst case is assumed, setting all the risk variables in the model to the riskiest level that Farmer Mac will generally purchase. Farmer Mac will generally not buy loans in excess of $1 million, but in certain circumstances, loan balances can be as high as $10 million in current dollars. In general, Farmer Mac will not buy a loan with LTV greater than 0.70.[13] Similar restrictions apply to Debt/Assets (maximum generally 0.5) and DSCR (maximum allowed of 1.25). Farmer Mac will purchase a loan that exceeds the standard on one underwriting variable, but not on all, and not without compensating factors. No loan would be purchased with both an LTV of more than 0.75 and a Debt/Assets of more than 0.5. In their 2003 annual report Farmer Mac indicates that only about 3% of their credit losses are from loans that exceeded 70% initial LTV. In testimony before the House Agriculture Committee, Farmer Mac's CEO indicated that their weighted average LTV was under 50% (Farmer Mac 2004). For these simulations initial LTV's are set to 0.75, and Debt/Asset ratios are set to 0.5, the Debt Service Coverage Ratio is set to 1.25, and the loan size is set to $11 million (1997 dollars). These values should produce a loan at the extreme of Farmer Mac's risk tolerance.

For the first comparison, all loans are assumed to be in their peak foreclosure year, the 6th year after origination. The credit risk model is used to forecast credit losses with the above worst case risk parameters as described above, for various levels of the price-shock variable, always assuming that the price shock occurs in the first year of loan life, the year with the biggest impact on credit loss. The Lifetime Foreclosure column is the probability of a loan going to foreclosure and resulting in a credit loss, calculated from the logistic regression. The Annual Foreclosure column multiplies the previous column by 0.168, to reflect the fraction of all foreclosures foreclosing in the 6th (peak) policy year, calculated from the Beta distribution. The Credit Loss columns multiply the respective Foreclosure columns by 0.209, to reflect the loss severity rate assumed in the model. The results are as follows:

Table 2 – Foreclosure Rates and Credit Loss Rates Predicted by RBC Model

(all figures in percent)

Price Shock	Lifetime Foreclosure	Annual Foreclosure	Lifetime Credit Loss	Annual Credit Loss
0	0.127	0.021	0.027	0.004
-2	0.250	0.042	0.052	0.009
-4	0.491	0.082	0.103	0.017
-6	0.961	0.161	0.201	0.034
-8	1.872	0.314	0.391	0.066
-10	3.618	0.608	0.756	0.127
-12	6.876	1.155	1.437	0.241
-14	12.683	2.131	2.651	0.445
-16	22.223	3.733	4.645	0.780

Price Shock is the value of the predictor variable, maximum land price decline. Lifetime Foreclosure is the predicted probability from the logistic regression using that value of Price Shock. Annual foreclosure is the Lifetime Foreclosure times the fraction of loans foreclosing in the peak year (0.168), estimated from the Beta distribution. Lifetime Credit Loss is the Lifetime Foreclosure Probability times loss severity (0.209), and Annual Credit Loss is the Annual Foreclosure Rate times loss severity.

These results may be compared with credit losses reported in Farmer Mac's 2003 annual report. Annual charges for its $5.5 billion portfolio of loans and guarantees are reported to be just over $5 million, for a charge-off rate of about 9 basis points. As Farmer Mac reports about $1 billion in new purchases for the year, (which are too recent to have produced credit losses and lie outside the scope of the 6th policy year worst case assumption), a fairer estimate is about an 11 basis point annual charge-off rate for seasoned loans. According to USDA NASS data, the worst price shock between 1994 and 2003 was in North Dakota, and was a fall of one-quarter of one percent. None of the western states in which Farmer Mac's loans are concentrated experienced any fall in farm real estate prices between 1994 and 2003. For the U.S. as a whole, USDA estimates a fairly steady increase in farmland real estate prices of about 5 per cent per year. Therefore, a conservative choice for an out of sample prediction is the first row, with a price shock of zero. Farmer Mac's annual charge off rate is more than 20 times the rate predicted by the credit risk model at a price-shock of zero, and is consistent with a price shock of between 8 and 10 per cent. To the extent that Farmer Mac's portfolio consists of better credit risks than those assumed in this worst-case analysis, the corresponding price-shock would need to be even more adverse.

Farmer Mac's 2003 annual report also gives cumulative net loss rates by cohort. Cumulative loss rates are as follows: for 1996, 0.35%, for 1997, 0.46%, for 1998 0.24%, for 1999 0.13%, and for 2000 0.23%. These 4 to 8 year loss rates all exceed the predicted lifetime loss rate for a price-shock of zero, and even exceed the predicted lifetime loss rates for a price shock of –4%. Using the Beta distribution estimates, it can be shown that a lifetime loss rate should be about 6 times the 4 year loss rate, and be about 50% larger than the 8 year loss rate. Extrapolating the cumulative loss rates to lifetime rates via the credit risk model's Beta distribution yields predictions consistent with shocks between -8% and -12%. Again, to the extent that Farmer Mac's portfolio consists of loans safer than the worst-case calculations given here, the corresponding price shock consistent with these net loss rates would be correspondingly more adverse.

The credit risk model substantially underpredicts risk at the safe end of the price-shock spectrum. As the price-shock foreclosure relationship is estimated off of two clusters of loans, those with shocks near –3% and those with shocks near –17%, it's inability to predict credit losses in the small shock region calls into question its ability to accurately estimate the slope of the price-shock foreclosure relationship, hence the reliability of the credit risk model's extrapolation[14] to a price shock of –24%. The consequences for the risk-based capital standard are unclear. If this result is driven by incomplete data at the safe end of the spectrum caused, for example, by right censoring of the loans not experiencing a price-shock, then the regression may overestimate the slope of the relationship, and project a higher capital requirement than is necessary. However, if the underprediction is driven by an undercount of credit risk in all ranges of the estimation sample, caused for example by zero credit loss foreclosures actually having credit losses, or estimation sample loans prepaying before experiencing the competing risk of default, the model may project a risk based capital requirement that is too small.

Conclusions

FCA's credit risk model, as currently implemented, has several limitations, both with respect to the data used in the analysis, and the structure of the equations in the econometric analysis of that data. The impact of these limitations is not clear. Some, such as the understatement of severities, probably lead to an underestimate of the credit risk on agricultural mortgages (unless there is a strong offsetting correlation between understatement and price shocks that flattened the price-shock foreclosure relationship). Others, such as the endogeneity of the land price shock variable, probably lead to an overestimate of credit risk.

A good estimate of the credit risk on these mortgages is important for managing the taxpayer's exposure to the implicit guarantee of the debt on this GSE. It is important to get both the required level of capital right, and to correctly estimate the marginal contribution to risk of different variables that may influence risk. For example, if large embedded prepayment penalties (such loans are common in a falling interest rate environment) increase risk, but the regulatory model in use does not capture the increased risk, the regulated institution has a regulatory incentive to purchase seasoned loans with large embedded penalties rather than newly originated loans with smaller penalties.

It is prudent to improve both the credit risk model and the underlying data. The data could be improved via updating the dataset with post 1992 terminations on the loans in the current model, and perhaps with post 1992 originations and terminations. Further refinements may be more speculative in nature, but it would be worth exploring many possibilities. For example, it may be possible to create an estimated prepayment date from the payment history file, so that competing risks and conditional default rates can be modeled. If not, aggregate FCS data or data from other FCS institutions may be used to produce an estimate of annual prepayment probabilities, so that loans known to have terminated by the end of 1992 could be probabilistically censored in earlier years. There may be other sources of data on actual economic losses, versus estimated accounting losses. Perhaps a random sample of foreclosed loans could be compared to actual county records of sales prices and dates for the REO collateral. Initial LTV's in the Texas data can be adjusted for stock purchase requirements, and the regulator could collect sufficient data to measure stock purchase, compensating balances, etc., so that the adjusted figures can be used to project credit risk on the loans in Farmer Mac's portfolio. It may be possible to use data on dates and geographic locations to tie together the old and new loans that result from restructuring, so that their performance can be modeled jointly.

The credit risk model could itself be improved in several dimensions. A hazard rate framework could solve the endogeneity of the key independent variable, land price shock. If the data permit, a competing risks framework could be adopted to reflect the substantial differences in prepayment incentives between the loans in the analysis dataset and the loans currently originated. Finally, more variables could be incorporated, such as the term of the mortgage, updated equity estimates stemming from amortization and land price changes, and erosions of equity stemming from prepayment penalties.

The risk-based capital constraint on Farmer Mac is not currently binding. The leverage capital requirement substantially exceeds the risk-based capital requirement. Thus, there is time to improve the modeling of the risk on agricultural loans before risk-based capital becomes a binding constraint.

As the Basel II regulatory regime comes into force, more and more institutions will be building their own credit risk models for regulatory capital purposes. Financial regulators need to establish criteria for acceptable credit risk models, and will also need to develop the capacity to identify flawed models before they are deployed.

References

Bank for International Settlements, "Principles for the Management of Credit Risk," July 2001, (BIS, Basel, Switzerland).

Barry, P. DeVuyst, Lins, Miller, and Sherrick 1993. "Stress Study of Agricultural Real Estate Loans," Center for Farm and Rural Business Finance, U. of Illinois, report to the Farm Credit Administration.

Capozza, D., Kazarian, D., and Thomson, T. 1998. "The Conditional Probability of Mortgage Default" Real Estate Economics V. 26 #3 pp. 359-389

Deng, Y., Quigley, and VanOrder 2000. "Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options," Econometrica, 68 (2), 275-307.

Farmer Mac "2003 Annual Report" prepared by Farmer Mac, Washington, DC.

Foster, C. and VanOrder, R 1985. "FHA Mortgage Terminations: A Prelude to Rational Mortgage Pricing," American Real Estate and Urban Economics Journal, 13 (3), pp 273-291

Kelly, A. 1998. "Capitalization of Above Market Financing: Condos and Coops" Jrnl. of Real Estate Research, V.15 #2 pp. 163-176

Kelly, A. and Slawson, V. C. 2001 "Time-Varying Prepayment Penalties and the Enhanced Value of Delay" Jrnl. Of Real Estate Finance and Economics, V23 #2, September 2001

Kennedy, P. 1987. *A Guide to Econometrics*, 2nd edition, (MIT Press, Cambridge, MA).

OFHEO 1993, Risk-Based Capital Regulation: Second Notice of Proposed Rule Making, Federal Register V64 #70, April 13 1999.

PriceWaterhouseCoopers 1997. "An Actuarial Review for Fiscal Year 1996 of the FHA Mutual Mortgage Insurance Fund." Report to HUD.

US GAO 1985. "Preliminary Analysis of the Financial Condition of the Farm Credit System," GAO/GGD-86-13-BR

US GAO 1993. "Homeownership: Appropriations Made to Finance VA's Housing Program May be Overestimated," GAO/RCED-93-173

US GAO 1996. "Mortgage Financing: FHA Has Achieved Its Home Mortgage Capital Reserve Target," GAO/RCED-96-50

US GAO 2003. "Farmer Mac: Some Progress Made, But Greater Attention to Risk Management, Mission, and Corporate Governance is Needed," GAO-04-116.

US GAO 2004. "Small Business Administration: Model for 7(a) Program Subsidy Had Reasonable Equations, but Inadequate Documentation Hampered External Reviews," GAO/04-9, 2004

U.S. House, Committee on Agriculture, *Testimony of Henry Edelman, CEO of Farmer Mac, to the House Committee on Agriculture* 108th Congress, 2nd session June 2, 2004.

VanDell, K. and Barnes, Hartzell, Kraft, and Wendt 1993. "Commercial Mortgage Defaults: Proportional Hazards Estimation Using Individual Loan Histories," AREUEA Journal V. 21 #4, pp. 451-480.

Yamaguchi, K. 1991. *Event History Analysis* (Sage Publications, Inc., Newbury Park, CA).

Footnotes

---

[1] Substantial background information on Farmer Mac, FCA, and FCA's risk based capital modeling can be found in GAO's report on Farmer Mac (US GAO 2003). This report and other post 1994 GAO reports are available at http://www.gao.gov.

[2] Capital must be at least 2.75 % of on-balance sheet assets plus 0.75% of off-balance sheet obligations, somewhat higher than the leverage requirements on Fannie Mae and Freddie Mac.

[3] For example, OFHEO uses a competing risks hazard rate framework for assessing residential mortgage risk (OFHEO 1999), GAO (US GAO 1993, 1996) and HUD (Price-Waterhouse Coopers 1997) have used a competing risks hazard rate framework for assessing residential mortgage insurance costs and SBA uses a competing risks hazard rate framework to estimate the costs of guaranteeing small business loans (US GAO 2004).

[4] "Unconditional probabilities are sufficient for unseasoned loans, while conditional probabilities are essential for seasoned loans." (Capozza, Kazarian, and Thomson, 1998).

[5] The CFR document which describes the model indicates that the consultants who built the model did not have access to "prepayment information." However, a previous paper (Barry, et al 1993) indicates that they did have access to a file of loan payment information, but did not use it. For a loan with a terminated status code, assuming a prepayment near the date at which regular payments cease should be a reasonable approximation.

[6] The estimation of the model used USDA ERS Statistical Series 86010, Farm Real Estate Values, which was superseded by a USDA NASS series in 1996. The "as of" date varied over time, sometimes January or February of the year, sometimes April 1st (ie, the first day of the 2nd quarter). This series can be downloaded at http://www.ers.usda.gov/data/archive/86010/

[7] About 20 (0.1% of the population) loans entered foreclosure before 1986 but did not result in a credit loss. These were assigned larger positive values for the land price change variable.)

[8] Lenders may have other forms of protection, such as requirements for compensating balances. The loans in Farmer Mac's portfolio may well have such extra protections. The point of the discussion is that mechanisms that can act like equity, such as mandatory stock purchases or compensating balances, should be included in the analysis.

[9] A suggested formula for the yield maintenance penalty can be found in Farmer Mac's Seller-Servicer Guidelines. The formula is similar to that contained in Fannie Mae's DUS manual, which is discussed in Kelly and Slawson (2001).

[10] I am not aware of any paper that explicitly treats prepayment penalties as a risk factor in mortgage foreclosures. However, many papers use the market value, as opposed to the book value, of the mortgage in a determination of an equity variable, which results in a similar calculation. For example, VanDell et al (1993) use market value in a model of commercial defaults. Kelly (1998) provides evidence that prepayment penalties erode equity. Capozza, Kazarian, and Thompson (1998) discusses the insufficiency of a book equity concept, and suggests additional variables be added to the regression to capture book-market differences if the book equity concept is used.

[11] The Bank for International Settlements defines credit risk as "the potential that a bank borrower or counterparty will fail to meet its obligations in accordance with agreed terms" (BIS 2001, p. 1). Since prepayment penalties are one of the "agreed terms" the BIS definition would seem to encompass a failure to pay these penalties. However, conversations with several risk

managers and regulators indicate that this risk is treated as market risk, since it stems from fluctuations in interest rates, when it is treated at all.

[12] For example, GAO's report on FHA (US GAO 1996) uses Heckman's CTM software to estimate Box-Cox baseline hazards simultaneously with risk factors such as equity and a Heckman-Singer non-parametric heterogeneity distribution. Deng, Quigley and VanOrder (2000) estimate a similar model on Freddie Mac data with non-parametric baseline hazards simultaneous with risk factors and Heckman-Singer non-parametric heterogeneity.

[13] Farmer Mac's underwriting guidelines generally specify a 0.70 LTV limit, but certain cash window loans allow LTV's up to 0.75. Part Time Farmer loans allow an LTV up to 0.85, but this program is very small and requires substantial private mortgage insurance.

[14] The model actually linearizes the functional form at the –17% price shock, but the linearized slope is still a function of the slope of the logistic regression, estimated between -3% and –17%.