Estimating Delinquency Migration and the Probability of Default from Aggregate Data

Jeffrey R. Stokes and Brent A. Gloy

Abstract

Defaulting on a mortgage represents the ultimate consequence of past decisions to delay payment. While many modeling approaches are available to estimate the probability of default, most if not all require account-level data. Further, past research has not attempted to estimate the probability that a current loan will transition among delinquency states prior to default. In this paper, we present an econometric approach that makes use of publicly available aggregate data for estimating the probability of delinquency and the probability of default. The results suggest the approach may have merit for monitoring bank performance as well as usefulness for banks' risk management efforts.

Key words: delinquency, Markov chain, maximum entropy, probability of default

Jeffrey R. Stokes is associate professor, Department of Agricultural Economics and Rural Sociology, Penn State University, and Brent A. Gloy is associate professor, Department of Applied Economics and Management, Cornell University.

Review coordinated by Calum G. Turvey.

Default and prepayment risk are a central concern for financial institutions for a variety of reasons including performance measurement, risk management, and regulatory compliance. These two risks have received considerable attention in the literature. For example, with respect to agricultural mortgages, Katchova and Barry (2005), Novak and LaDue (1994, 1997), and Sherrick, Barry, and Ellinger (2000), all concentrate on credit (i.e., default) risk, and Chhikara and Hanson (1993), Brinch and Stokes (2001), and Stokes and Brinch (2001) focus on prepayment risk in addition to credit risk.

While these studies and others like them are important for understanding credit and prepayment risk, the literature has much less to say about the estimation of the risk associated with mortgage delinquency. Although mortgage delinquency is not a sufficient condition to ensure default, it is a necessary condition for default, as loans will generally progress through various stages or states of delinquency prior to an ultimate default. In addition, there is a direct linkage between delinquency and credit and prepayment risk. Delinquent loans are more likely to default, while current (i.e., nondelinquent) loans are more likely to remain current or prepay.

In one of the first studies of mortgage delinquency, Green and von Furstenberg (1975) attribute the paucity of research on delinquency to a lack of relevant data appropriate to analyze the occurrence. Only a few other studies (such as Morton, 1975; Campbell and Dietrich, 1983; Teo, 2004; and Diaz-Serrano, 2005) examine delinquency risk specifically. Campbell and Dietrich (1983) approach the issue of delinquency from the perspective of a utility-maximizing borrower operating in a dynamic choice environment. At each point in time of the life of a mortgage, borrowers observe a set of state variables and choose one of four utility-maximizing actions: default, delay payment, prepayment, or continue to service the mortgage. Similarly, in their review of the mortgage literature, Quercia and Stegman (1992) describe typical borrower repayment models as depicting borrowers dynamically selecting one of these optimal actions to maximize their utility.

In reality, default may not be a decision at all, but rather the ultimate consequence of previous decisions to delay payment (i.e., delinquency) and subsequent failure to bring the loan back to current status by making up the payments. In other words, there may be something to gain by estimating the probability that loans transition among delinquency states.

One obvious drawback to estimating the probability of default and potentially the probability of delinquency using current methods is the need for micro-level data—that is, loan- or account-level data where the status (i.e., current, past-due, default, etc.) of the loan can be observed at discrete points in time. Unlike Gloy, LaDue, and Gunderson (2005), and Zech and Pederson (2004), applied researchers rarely have access to such micro-level data. This typically means that two choices are available to the applied researcher: namely, to not attempt any estimation at all, or to seek out a modeling approach that might be sensitive to available data.

Given the lack of research on mortgage delinquency in general and for agricultural loans specifically, we take the latter approach and model mortgage delinquency as a Markov chain using publicly available data. In our model, a loan can remain current, transition among various states of delinquency, or charge off (default) from period to period. The dollar volume of loans in each state of the chain is observable, but the transition of accounts among states of the Markov chain over time is not. Faced with this ill-posed problem, we use an entropy-based econometric technique to elicit the probability distribution of the transition probabilities.

Through our demonstration of the entropy approach, our contribution is to add to the number of ways to estimate the probability of default and also to the number of ways to estimate credit ratings migration since our approach can be applied in this setting as well. Last, the Markov chain framework for loan delinquency necessitates consideration of how to handle the bank's changing loan volume over time. When using the Markov model, it is typical to specify an ad hoc pool size from which the proportion of the observations in each state including the pool can be determined. These data are then used to estimate the transition probabilities. We take a more innovative approach and specify a model where the size of the bank's loan pool is estimated simultaneously with the transition probabilities. This pool size estimate gives an indication of how much loan volume is both realized and unrealized by the bank and is therefore also useful for bank performance measurement and risk management.

Delinquency Dynamics

Whenever a stochastic process can frequent only one of a finite number of potential states at discrete points in time, the discrete time Markov chain model is a potentially appropriate description of the stochastic process. More formally, a Markov process $\{X_t\}$ is a stochastic process with the property that, given the value of X_t, the values of X_s for $s > t$ are not influenced by the values of X_u for $u < t$ (Taylor and Karlin, 1994). This leads to the Markov property

$$\Pr\{X_{n+1} = j \mid X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i\} = \Pr\{X_{n+1} = j \mid X_n = i\} = p_{ij},$$

for all time points n and all states $i_0, \ldots, i_{n-1}, i, j$.

Since bank loans can be classified according to delinquency status at discrete points in time, the Markov chain model is one way to characterize the dynamics of bank loan delinquency. For example, the finite state space is comprised of a current state which shows how much of the bank's portfolio is made up of loans that are being paid back on schedule, various delinquency states which show how much of the bank's portfolio is past due and by how many days, and default which shows how much of the bank's portfolio charged off over the discrete time step between observations of the data.

The p_{ij} of the Markov chain are then the (stationary) probability of loans in one state transitioning to any of the other states over a discrete period of time. All Federal Deposit Insurance Corporation (FDIC) insured financial institutions report on the dollar value of loans that are current, 30–89 days past due, greater than 90 days past due, nonaccruing, and charged off during a quarter. Arranged into a matrix, **P**, defined by these states, a quarterly transition probability matrix can be represented as:

(1) $\mathbf{P} =$

$$\begin{bmatrix} p_{cu,cu} & p_{cu,30} & p_{cu,90} & p_{cu,na} & p_{cu,df} & p_{cu,ex} \\ p_{30,cu} & p_{30,30} & p_{30,90} & p_{30,na} & p_{30,df} & p_{30,ex} \\ p_{90,cu} & p_{90,30} & p_{90,90} & p_{90,na} & p_{90,df} & p_{90,ex} \\ p_{na,cu} & p_{na,30} & p_{na,90} & p_{na,na} & p_{na,df} & p_{na,ex} \\ p_{df,cu} & p_{df,30} & p_{df,90} & p_{df,na} & p_{df,df} & p_{df,ex} \\ p_{en,cu} & p_{en,30} & p_{en,90} & p_{en,na} & p_{en,df} & p_{en,ex} \end{bmatrix},$$

where the subscripts are defined as follows: *cu* (current), 30 (30–89 days past due), 90 (greater than 90 days past due), *na* (nonaccruing), *df* (default), *en* (entry), and *ex* (exit). Assuming quarterly data, one quarter is represented on the left side of the matrix with a subsequent quarter on the upper portion of the matrix whereby, for example, $p_{cu,cu}$ represents the probability of a current loan remaining current one quarter later. The entry and exit states are required to allow new loans to enter the bank's balance sheet and existing loans to exit via prepayment or maturity. In addition, some of the probabilities appearing in (1) must be restricted in value. These ideas are developed more fully later in the paper.

Estimating the probabilities in (1) is a complicated task for a variety of reasons. First, micro-level data documenting state transitions would typically be required for the estimation of the transition probabilities. Lee, Judge, and Takayama (1965) have shown that maximum-likelihood estimates of stationary transition probabilities are relatively easy to obtain when micro-level data are available. If micro-level data are unavailable or if the Markov chain is not stationary, estimation can be much more difficult. In the stationary case when micro-level data are unavailable, proportional data have been shown to be useful for the estimation of transition probabilities (Lee, Judge, and Zellner, 1970). However, it is often the case that the Markov chain problem is ill-posed since a short time series of data can easily cause the number of unknown parameters in the model to exceed the number of data points (Golan, Judge, and Miller, 1996), making the model underdetermined. This is easily the case in the context of estimating transition probabilities for the matrix in (1). With six states, 25 parameters (transition probabilities) would need to be estimated for a stationary Markov chain.[1]

Further complicating the estimation are the entry and exit states which allow new loans to enter and prepays or mature loans to exit the bank's portfolio. When using the Markov chain model, a pool state is often specified to allow for such entry and exit. However, when using proportional data, the size of the pool is

[1] A 6 × 6 square matrix has 36 probabilities. However, the probabilities in each row must sum to one and the default state is an absorbing state leaving at least 25 unique probabilities to be estimated. Other restrictions may further reduce the number of parameters to be estimated.

unknown and has implications for the estimation of the transition probabilities (Stanton and Kettunen, 1967). The next section describes the data and modeling approach used in this study and the unique way in which pool size is accommodated.

Empirical Model

Following the early work of Shannon (1948), Golan, Judge, and Miller (1996) propose a system for estimating the transition probabilities of a Markov chain based on maximizing the Shannon entropy function. In the simplest case, the Markov problem is cast as a pure inverse problem:

$$
\begin{aligned}
(2)\quad & \max -\sum_i \sum_j p_{ij} \ln(p_{ij}) \\
& \text{s.t.: } y_j(t+1) = \sum_i y_i(t) p_{ij} \quad \forall\, t, j, \\
& \sum_j p_{ij} = 1 \quad \forall\, i, \\
& 0 \le p_{ij} \le 1 \quad \forall\, i, j.
\end{aligned}
$$

In (2), the p_{ij} represent the stationary probability of transitioning from state i to state j over one time period. The objective function measures the Shannon entropy function which takes a maximum when the distribution of transition probabilities is uniform. The first set of constraints represents the Markov condition for proportional time-dated data $y_i(t)$ and $y_j(t+1)$,[2] and the second set of constraints are the row sum conditions associated with the Markov chain. The last set of constraints ensures the estimated probabilities are proper probabilities.

The concept of entropy in this context is relatively simple. Specifically, we seek to model all that is known and assume nothing about what is unknown. In other words, given a collection of data (information), choose a model which is consistent with all the facts, but which is otherwise as uniform as possible. The entropy of the distribution of transition probabilities is maximized in an effort to reduce information uncertainty.

The system presented in (2) is a pure inverse problem and is only appropriate if the data-generating process is Markov and, further, the data are observed without error. While the former assumption for delinquency is realistic at least as a working assumption, the latter is unrealistic. Point estimates like those resulting from a system such as (2) are likely less appealing than a range of probabilities from which estimation precision can be determined. Additionally, non-sample information may be available and desirable to incorporate into the estimation.

Golan, Judge, and Miller (1996) suggest using a specification such as that presented in (3) subject to (4), where the system (2) has been augmented to accommodate these deficiencies. The minimum cross-entropy formalism of the stationary Markov problem is constructed as:

$$
\begin{aligned}
(3)\quad \min_{\pi_{ijm},\,\omega_{ijm}} \boldsymbol{\Psi}(\boldsymbol{\pi}, \boldsymbol{\omega}) = & \sum_i \sum_j \sum_m \pi_{ijm} \ln\left(\frac{\pi_{ijm}}{\tilde{\pi}_{ijm}}\right) \\
& + \sum_t \sum_j \sum_m \omega_{tjm} \ln\left(\frac{\omega_{tjm}}{\tilde{\omega}_{tjm}}\right)
\end{aligned}
$$

subject to:

$$
\begin{aligned}
(4)\quad & y_j(t+1) - \sum_i y_i(t) \sum_m z_m \pi_{ijm} \\
& \qquad + \sum_m v_m \omega_{tjm} = 0 \quad \forall\, t, j, \\
& 1 - \sum_j \sum_m z_m \pi_{ijm} = 0 \quad \forall\, i, \\
& 1 - \sum_m \pi_{ijm} = 0 \quad \forall\, i, j, \\
& 1 - \sum_m \omega_{tjm} = 0 \quad \forall\, t, j, \\
& 0 \ge \pi_{ijm}, \omega_{tjm} \le 1.
\end{aligned}
$$

In (3) subject to (4), the sum product of a discrete distribution of estimated transition probabilities, π_{ijm}, and a parameter support vector, $\mathbf{z}$, determine the desired transition probabilities, p_{ij}, i.e.,

[2] Proportional data are assumed since they are all that is publicly available. Lee, Judge, and Zellner (1970) show how proportional data can be modeled as a Markov chain.

$$p_{ij} = \sum_m z_m \pi_{ijm}.$$

Additionally, the specification allows for the possibility that there is error in the Markov relation in (4) by including an error term expressed as the sum product of an error support vector, $\mathbf{v}$, and errors, ω_{ijm}, i.e.,

$$\varepsilon_{ij} = \sum_m v_m \omega_{ijm}.$$

The remaining equations ensure that the estimated probabilities are proper probabilities. Non-sample information is introduced through the prior probability distributions for the probabilities and errors and are expressed in the objective function as $\tilde{\pi}_{ijm}$ and $\tilde{\omega}_{ijm}$. Unless prior distributions are explicitly specified, a uniform prior is implicitly assumed by the specification in (3).

Data

To estimate the transition probabilities, a time series of actual bank data showing the proportion of agricultural loans in various delinquency categories is required. This study utilizes quarterly observations on the dollar value of agricultural real estate loans in current, past due 30–89 days, past due more than 90 days, nonaccrual, and charge-off states. The data were collected for Pinnacle Bank in Pappilion, Nebraska, which until the most recent quarter was the largest agricultural bank by loan volume.

Data at this level of detail are available for 18 quarters beginning in March 2001 and ending June 2005. Before March of 2001, banks reported the amount of loans in the past-due more than 90 days, nonaccrual, and charge-off states, but the current and 30–89 days past-due states were combined for confidentiality purposes. The relatively short time period (4½ years) makes a stationary modeling approach justifiable at least as an initial modeling. In addition, the short time series make the entropy approach imperative since there are more parameters to estimate than data points, as discussed below.

The FDIC data were used to construct shares of agricultural real estate loans in each state for each quarter. In the process of creating the shares, it is necessary to modify the charge-off data because charge-off is an absorbing state. When viewed as an absorbing state, reported charge-off data are flow data, whereas the amount of loans in the other states are stocks. Therefore, for consistency with the other data in the model, the charge-off data must be converted to a stock. This was accomplished by accumulating the dollar value of loans flowing into the charge-off state over the sample period. While this adjustment is necessary for the estimation of transition probabilities for states communicating with charge-off, transition probabilities from the charge-off state to other states are necessarily zero. In other words, once a loan enters the charge-off state it cannot leave that state.

New loan volume and loan volume that leaves the bank's portfolio through means other than default (either prepayment or through maturity) must be accounted for as well, since failure to do so would result in the implicit assumption that loan volume expansion and contraction occurs proportionally to the states modeled.[3] However, from quarter to quarter, the data available only show the net change (increase or decrease) in the size of the bank's agricultural loan portfolio.

Given this aggregated feature of the data, an additional state representing a pool of new and retired loan volume is necessary. Any new loans come into the system from the pool and any repaid loans transition from the bank's balance sheet to the pool. The difficulty here is that the size of the pool represents the magnitude of Pinnacle's realized and unrealized loan volume, an unknown dollar value that affects the magnitude of the remaining shares. The minimum size of the pool can

[3] For example, by not allowing entry and exit, a 10% contraction in overall loan volume from one quarter to the next would presuppose an unrealistic 10% contraction in current loan volume, a 10% contraction in 30–89 days past-due loan volume, etc.

be inferred directly from the data as the maximum loan volume actually experienced by the bank over the sample period. The bank's portfolio expands over a quarter whenever loan volume increases. This expansion causes the pool of loan volume to shrink, as was typically the case for Pinnacle Bank over the sample period. The transition probabilities associated with the pool indicate the likelihood that loans in the bank's portfolio transition to the pool (exit) by means other than default, or enter the bank's portfolio from the pool.

While knowing the minimum pool size is useful information, imposing it on the share data is ad hoc because the pool represents an unknown loan volume which may be considerably larger than minimum observed loan volume depending on the bank's operating and competitive situation. More importantly, Stanton and Kettunen (1967) have shown that the number of potential entrants to a Markov chain system affect the estimation of all the transition probabilities and the equilibrium distribution of the system. Consequently, a significant extension of the Markov model is the simultaneous estimation of the size of the bank's loan pool and the transition probabilities. Such an approach thereby negates the need for an assumed pool size as in all previous structural change studies making use of the Markov model. Additionally, the estimate provides a bank with information about the size of its loan pool which can be compared to its realized or observed loan volume to gauge the extent to which resources are being allocated properly to enhance competitiveness.

As noted above, the difficulty in adding this feature to the model is that the magnitude of the bank's loan pool directly affects the share data, $y_i(t)$ from which estimates of the transition probabilities originate. Let $i = N$ represent the pool state and $Y_i(t)$ represent dollars of loan volume in state i at time t. Let $\gamma(t)$ represent total loan volume in states other than N so that

$$(5) \quad \gamma(t) = \sum_{i \neq N} Y_i(t) \quad \forall\, t.$$

Next, let $\eta(t)$ represent the net change in loan volume from one quarter to the next whereby

$$(6) \quad \eta(t) = \gamma(t+1) - \gamma(t) \quad \forall\, t < T \quad \text{and} \quad -\infty \leq \eta(t) \leq \infty.$$

Letting $\mathbf{d}$ be a parameter support vector for the unknown loan volumes, the magnitude of the pool at time $t = 1$ can be determined using (5) so that

$$(7) \quad Y_N(t) = \sum_m \beta_m d_m - \gamma(t),$$

while for $t > 1$ we have

$$(8) \quad Y_N(t) = Y_N(t-1) - \eta(t),$$

where β_m are probabilities associated with each parameter support value with

$$\sum_m \beta_m = 1.$$

To complete the estimation, let $\tau(t)$ represent the sum of observed and unobserved loan volume so that

$$(9) \quad \tau(t) = \gamma(t) + Y_N(t) \quad \forall\, t.$$

The proportion of loans in each state required by the Markov relation in (4) are then determined as

$$(10) \quad y_i(t) = \frac{Y_i(t)}{\tau(t)} \quad \forall\, t, i.$$

Augmenting the objective function in (3) to formalize the cross-entropy estimation of transition probabilities and loan volume results in the new objective function:

$$(11) \quad \mathbf{\Psi}(\boldsymbol{\pi}, \boldsymbol{\omega}, \boldsymbol{\beta}) = \sum_i \sum_j \sum_m \pi_{ijm} \ln\left(\frac{\pi_{ijm}}{\tilde{\pi}_{ijm}}\right) + \sum_t \sum_j \sum_m \omega_{tjm} \ln\left(\frac{\omega_{tjm}}{\tilde{\omega}_{tjm}}\right) + \sum_m \beta_m \ln\left(\frac{\beta_m}{\tilde{\beta}_m}\right),$$

with the new third term in (11) reflecting the uncertainty in the size of the loan pool facing the bank. In (11), $\tilde{\boldsymbol{\beta}}$ is a vector of prior probabilities for loan volume. Notice that the model chooses probabilities for each loan support value. The sum product of these probabilities and support values is then used via equation (7) to determine the proportion of loan volume in each of the remaining states. This feature of the model is perhaps the most compelling reason for the entropy specification in that the number of unknowns far exceeds the number of data points. Equation (11) is minimized subject to the constraints in (4)–(10).

Finally, transition probabilities from the pool to certain states must be restricted to zero when using quarterly data. For example, it is impossible for a loan originated in one quarter to transition to the nonaccrual state one quarter later. Therefore, the transition probabilities from the pool and current states to the three states of (*a*) greater than 90 days past due, (*b*) nonaccrual, and (*c*) charge-off are restricted to zero. Also restricted to zero are transition probabilities for movement from the 30–89 days past-due (more than 90 days past-due) state to the nonaccrual and charge-off (charge-off) states. Last, transitioning from the nonaccrual state back to the 30–89 days past-due and more than 90 days past-due states in one quarter is also precluded.

Since the transition probabilities are bounded between zero and one, a logical choice for the probability parameter support vector would be discrete points in the unit interval, such as $\mathbf{z}$ = [0 ¼ ½ ¾ 1]. Similarly, the error support vector is specified as $\mathbf{v}$ = [−1 −½ 0 ½ 1] which imposes a symmetric error distribution with values consistent with the magnitude of error possible given the size of the probabilities to be estimated. That is, the most we can misestimate a transition probability is by a magnitude of one and the error support chosen captures this possibility. Finally, the parameter support vector for potential loan demand is specified (in millions) as $\mathbf{d}$ = [\$100 \$150 \$200 \$250 \$300].[4] Prior probabilities for the transition probabilities ($\tilde{\pi}_{ijm}$), errors ($\tilde{e}_{ijm}$), and potential loan demand ($\tilde{\boldsymbol{\beta}}$) were all assumed to be uniform.

Results

With six states and five parameter supports for each state, it is not practical to present the results from the full estimation. In addition, transition probabilities are probably best thought of in an annual context and the estimation, having used quarterly data, results in quarterly transition probabilities. To circumvent these issues, presented in Table 1 is a matrix of annualized stationary transition probabilities for Pinnacle's loan portfolio. These probabilities were determined by first estimating the π_{ijm} by minimizing (11) subject to (4)–(10), then recovering quarterly probabilities via

$$p_{ij} = \sum_m z_m \pi_{ijm},$$

and finally raising the resulting quarterly matrix to the 4th power to put the estimates in annual terms.

The resulting matrix contains annual estimates of the probability of loans in Pinnacle's portfolio transitioning to various stages of delinquency, default, and exit (prepayment or maturity) over the course of one year. For example, a loan that is current has about a 60.1% chance of remaining current next year, a 6.5% chance of becoming 30–89 days past due, a 2.0% chance of becoming more than 90 days past due, a negligible chance (due to rounding) of entering nonaccrual status, and a zero probability of defaulting. The data also suggest that a current loan has a 31.4% chance of either prepaying or maturing in the next year.

---

[4] Pinnacle's maximum loan volume over the sample period is suggestive of a minimum pool size of \$151 million. The parameter support vector specified contains this value but allows for a larger loan volume if the data suggest this is the case.

Table 1. Annualized Estimates of Transition Probabilities and (Normalized Entropy Measures) for Pinnacle Bank

	Current	30–89 Days Past Due	90 or More Days Past Due	Non-accrual	Charge-Off (default)	Exit (prepayment or mature)
Current	60.08% (0.9894)	6.46% (0.0029)	2.03% (0.0000)	0.00% (0.0000)	0.00% (0.0000)	31.43% (0.9854)
30–89 Days Past Due	51.89% (0.8403)	9.71% (0.8057)	5.92% (0.8343)	3.39% (0.0000)	1.24% (0.0000)	27.86% (0.8574)
90 or More Days Past Due	49.65% (0.7601)	8.81% (0.7441)	4.80% (0.7550)	4.00% (0.7551)	7.20% (0.0000)	25.55% (0.7607)
Nonaccrual	39.90% (0.8413)	4.29% (0.0000)	1.16% (0.0000)	1.54% (0.8330)	32.55% (0.8328)	20.57% (0.8331)
Charge-Off (default)	0.00% (0.0000)	0.00% (0.0000)	0.00% (0.0000)	0.00% (0.0000)	100.00% (0.0000)	0.00% (0.0000)
Entry (new loan)	63.50% (0.7298)	9.43% (0.7298)	2.00% (0.0000)	0.93% (0.0000)	0.00% (0.0000)	24.14% (0.0000)

This result appears high but is not completely unanticipated, given the period of time covering the sample data was largely a time of falling interest rates in which many prepayments occurred.

The fact that the estimated probability of default for a current loan is zero is potentially problematic since no matter how high quality a loan is, there is always at least a small probability of default. However, the result is consistent with many past studies showing the same result for highly rated bonds. For the case of agricultural loans, estimates by Gloy, LaDue, and Gunderson (2005) and Behrens and Pederson (2005) indicate that the likelihood of a high quality credit, such as a current loan, transitioning to default is very nearly zero.

Below each transition probability estimate is a measure referred to as the normalized Shannon entropy measure. The statistic measures the information content in the estimates and is bounded by zero and one. Maximum entropy (or minimum cross-entropy) uncertainty is consistent with values for normalized entropy equal to one, while values closer to zero reflect less entropy uncertainty. The restrictions discussed above necessarily imply a zero value for the restricted parameters (e.g., the probability of transitioning from entry to past-due 90 or more days). However, most parameters have normalized entropy measures consistent with less information uncertainty. This is especially true for the probability of default from all states except nonaccrual. Obvious exceptions are current-to-current and current-to-exit transition probabilities. The overall system normalized entropy is 0.6893, which is reasonably consistent with less information uncertainty since it is fairly far from one.

Considering the remainder of the matrix, the probability of default rises as delinquency increases, and culminates in a 32.6% probability of default once a loan enters the nonaccrual state. This result makes sense because, as noted above, the decision to delay payment (i.e., become delinquent) is a necessary condition for default. However, it is important to point out that according to the estimation, Pinnacle's delinquent loans are most likely to transition back to the current state or exit the portfolio (by means other than default) in a year's time. This is likely because problem loans are identified and dealt with via a workout or prepayment rather than allowing them to progress through delinquency to default.

The results also suggest that new loans are most likely to be current a year later (about 64%), while a fair percentage of new loans can be expected to be retired (about 24%) via prepayment. The former result is obviously what the bank wants, whereas the latter is most likely a function of the time period over which the estimation was made (i.e., a period of generally declining interest rates). In any event, Pinnacle's success as a large agricultural lender can in some way be linked to its ability to generate new loans having a high probability of remaining current or prepaying.

The econometric estimation also produces estimates of the size of the bank's pool of loans, reported in Table 2. As shown, the probabilities are not quite uniform (as was the prior probability distribution). Multiplying the parameter support values times the probability estimates and summing gives an estimate of Pinnacle's observed and unobserved loan volume which equals $187 million. The estimate appears reasonable considering Pinnacle's maximum loan volume, achieved in the most recent quarter for which data were available, was $151 million. This result suggests that about $36 million in loan volume was unrealized by Pinnacle over the sample period. However, the normalized entropy is 0.9892, indicating a high degree of uncertainty about the probability distribution of this parameter. Even so, the magnitude of the pool of loans was endogenously determined by the data rather than exogenously by some ad hoc means as in past studies of structural change.

Table 2. Potential Demand Probability Estimates for Pinnacle Bank

d_m ($ millions)	β_m
$100	25.61%
$150	22.43%
$200	19.65%
$250	17.22%
$300	15.09%

Summary and Conclusions

Because delinquency is a precursor to default, estimating the probability that loans progress through delinquency states is important for the performance measurement and risk management of modern financial institutions. By viewing the bank's portfolio of loans as a Markov chain, it is demonstrated here how maximum entropy can be used to estimate the probability of delinquency which also contains the probability of default and the probability of new loan entry and old loan exit. Most importantly, the estimates can be obtained from publicly available data that are aggregate in nature. Such data are usually all that is available for researchers and investors alike. The model and estimation procedure presented in this study demonstrate how to obtain estimates with aggregated data by employing entropy econometrics.

While numerous techniques for estimating the probability of default have been proposed in the literature, the research outlined in this paper adds to the number of ways by suggesting a minimum cross-entropy formulation of the Markov chain model. The model presented has applicability when micro-level data are available, and therefore likely has usefulness for estimating credit ratings migration matrices and the probability of default, both of which are important for regulatory compliance or Basel consistency.

In addition, a nonstationary matrix of transition probabilities can also be accommodated with the approach we have outlined by making some straightforward adjustments to the estimation procedure. To accomplish this, equation (11) would be minimized subject to equations (4)–(10) for each transition of the data. Employing the 18 quarters of data used for the stationary estimation, nonstationary estimates could be obtained, resulting in 17 transition probability matrices—one for each transition of the data. Further, macroeconomic variables hypothesized to influence the transition probabilities could also be added to the estimation in both stationary and nonstationary settings. These extensions are left for future research.

References

Behrens, A., and G. Pederson. "Credit Risk Migration Patterns of Agricultural Loans." Paper presented at the Regional Research Committee NC-1014 annual meeting, Minneapolis, MN, 2005.

Brinch, B., and J. Stokes. "Quantifying the Impact of Farmer Mac Prepayment Penalties." *Agr. Fin. Rev.* 61(2001): 141–166.

Campbell, T., and J. Dietrich. "The Determinant of Default on Insured Conventional Residential Loans." *J. Finance* 38(1983):1569–1581.

Chhikara, R., and S. Hanson. "Competitive Pricing of Farmer Mac's Mortgage-Backed Securities and Its Future Viability: A Contingent Claims Analysis Approach." *Rev. Agr. Econ.* 15(1993):547–566.

Diaz-Serrano, L. "Income Volatility and Residential Mortgage Delinquency Across the EU." *J. Housing Econ.* 14(2005):153–177.

Gloy, B., E. LaDue, and M. Gunderson. "Credit Risk Migration and Downgrades Experienced by Agricultural Lenders." *Agr. Fin. Rev.* 65(2005):1–16.

Golan, A., G. Judge, and D. Miller. *Maximum Entropy Econometrics: Robust Estimation with Limited Data.* New York: John Wiley and Sons, 1996.

Green, R., and G. von Furstenberg. "The Effects of Race and Age of Housing on Mortgage Delinquency Risk." *Urban Studies* 12(1975):85–89.

Katchova, A. L., and P. J. Barry. "Credit Risk Models and Agricultural Lending." *Amer. J. Agr. Econ.* 87(2005): 194–205.

Lee, T., G. Judge, and T. Takayama. "On Estimating the Transition Probabilities of a Markov Process." *J. Farm Econ.* 47(1965):742–762.

Lee, T., G. Judge, and A. Zellner. *Estimating the Parameters of the Markov Probability Model from Aggregate Time-Series Data.* Amsterdam: North-Holland Pub. Co., 1970.

Morton, T. "A Discriminate Function Analysis of Residential Mortgage Delinquency and Foreclosure." *J. Amer. Real Estate and Urban Econ. Assoc.* 3(1975):73–90.

Novak, M., and E. LaDue "An Analysis of Multiperiod Agricultural Credit Evaluation Models for New York Dairy Farms." *Agr. Fin. Rev.* 54(1994): 55–65.

———. "Stabilizing and Extending Qualitative and Quantitative Indicators of Creditworthiness in Agricultural Credit Scoring Models." *Agr. Fin. Rev.* 57(1997):39–52.

Quercia, R., and M. Stegman. "Residential Mortgage Default: A Review of the Literature." *J. Housing Research* 3(1992):341–379.

Shannon, C. "A Mathematical Theory of Communication." *Bell System Technical J.* 24(1948):379–423.

Sherrick, B., P. Barry, and P. Ellinger. "Valuation of Credit Risk in Agricultural Mortgages." *Amer. J. Agr. Econ.* 82(2000):71–81.

Stanton, B., and L. Kettunen. "Potential Entrants and Projections in Markov Process Analysis." *J. Farm Econ.* 49(1967):633–642.

Stokes, J., and B. Brinch. "Valuing Agricultural Mortgage-Backed Securities." *J. Agr. and Appl. Econ.* 33(2001):493–511.

Taylor, H., and S. Karlin. *An Introduction to Stochastic Modeling.* New York: Academic Press, 1994.

Teo, A. "Delinquency Risk in Residential ARMs: A Hazard Function Approach." *J. Real Estate Portfolio Mgmt.* 10(2004):243–258.

Zech, L., and G. Pederson. "Application of Credit Risk Models to Agricultural Lending." *Agr. Fin. Rev.* 64(2004):91–106.