RATIO VS REGRESSION ESTIMATORS IN A LARGE SCALE SURVEY OF S&L's

Charles D. Cowan, Resolution Trust Corporation
5653 Kathryn St., Alexandria, VA 22303

KEY WORDS: Longitudinal, Stratification, Financial

Introduction

In June of 1991, the Resolution Trust Corporation (RTC) initiated a survey of the institutions it has taken into receivership since the RTC was founded in 1989. The purpose of the survey, called the Estimated Cash Recovery (ECR) Survey, was to provide quarterly estimates to Congress, the Administration, and the public in general of the amount of recovery expected from the sale of assets from failed Savings and Loan Institutions (S&Ls). An asset for an S&L is a loan made by the S&L for one of a variety of purposes (like a commercial loan or a construction loan), or a type of property that the S&L held or had received as collateral for a loan. There were 19 types of assets covered by the survey; the list of assets covered is found in Table 1.

The amount of recovery expected is the total recovery in dollars, summed over the 19 asset categories. Of almost equal interest is the recovery rate, defined as the total amount expected to be recovered divided by the current book value of the assets. The current book value is defined to be the original amount of loan, minus payment received on the loan. On performing loans, loans which are still being paid off by the person or business who took out the loan, the book value continually declines as the principal is paid off.

There are several complicating factors that made this an interesting and complex sample design. First of all, since there is some interest in the recovery rates for each of the asset categories, one has to trade off the design for an estimate of the total with the design to estimate each asset category total separately. Optimizing the sample to answer each of these goals would result in two very different sample designs.

Second, the population for this survey is changing very rapidly. Each quarter a number of institutions are transferred by the Office of Thrift Supervision to the RTC for resolution. At time of resolution, the depositors at each institution are paid off and the assets of the institution are taken over by the RTC. So there is a steady inflow of institutions and assets into the population. At the same time, assets from institutions previously taken over by the RTC are being sold. Assets in different categories are sold at very different rates; e.g. loans on single family dwelling units are sold much more rapidly than construction loans. So there is also a steady outflow of assets, and the flow out is at a very different rate than the inflow of assets.

Finally, there is some information available from the accounting ledgers of the RTC

that can be used in estimation to help reduce the variance of the estimates, using either a ratio or regression estimator. The recovery rate mentioned earlier is a good example of the type of ratio estimator that would be of interest. However, because of the nature of the processes that relate the recovery to the original book value of the asset, the recovery rate should be treated as bounded below by zero and above by unity. We can use the relationship between expected recovery value and book value to reduce the variance of the estimates (since book value is known for all population members), and achieve further reductions by bounding the estimates so that the range of these estimates is between zero and one.

This paper presents methods used to design and implement the ECR with some discussion of the problems encountered with the conduct of the survey. The paper concludes with some preliminary results from the survey collected over the first four quarters.

Methods

A sample of institutions and assets within institutions is selected, using the sampling methodology described below. Assets within institutions are valued when first sampled, and then rotate in and out of sample according to a prespecified schedule (also described in the next section). Chart 1 describes this process. Once the sample selected, the list of assets to be valued is given to a contracting firm of accountants who estimate how long it will take to sell the asset and the amount the RTC can expect to receive. The accountants also determine the flow of operating income for the asset or property, and the direct expenses incurred in the management of the asset.

The information on the expected flow of operating income, payments on the loan, and direct expenses are recorded on a quarterly basis for the two years following the date of data collection, and on an annual basis for years three, four, and five after the date of collection. The quarterly information is used to permit "rolling over" the estimates in subsequent quarters when the sampled institutions have rotated out of sample.

The other major part of the data collection is determining if a sampled asset is sold in subsequent quarters. If an asset is sold, it is no longer part of the survey process because the RTC has already realized the recovery. The survey's purpose is to estimate the future recovery to be realized - for sold assets we know exactly what has already been recovered so it is not necessary to sample these assets. The specialists who value assets in the field are also responsible for reporting the sale value of sold assets.

Sample Design

The plan was to develop a sample of between 5000 and 6000 assets nationally to be valued by the accountants. To make this as efficient as possible, both from a variance standpoint and a cost standpoint, the sample was designed to be a stratified multistage cluster sample. In addition, each asset category was considered

to be a separate stratum in the second stage of selection.

Stratification was conducted at the first stage by constructing a two way table of institutions, a table that would change at each quarter as the number of resolved institutions grew. The stratification variables were region, with categories East, North Central, Southwest, and West - defined in Chart 2 - and initial size of institution. Categories "Less than $100 million", "between $100 and $500 million", and "Greater than $500 million".

Strata were not balanced in terms of population size of institutions because there are many more small institutions than large. The strata are also not balanced in terms of total size (defined by book value) and number of assets, but the imbalance goes in the other direction, with the four strata with initial size greater than $500 million holding up more than half the total assets.

There were several conflicting factors to be accounted for in the stratification and estimation:

1) the distribution of the assets was heavily skewed to the larger institutions,
2) the number of institutions was heavily skewed to the smaller institutions,
3) the primary goal of the survey was to produce a single national estimate of recovery,
4) not all institutions have all types of assets, so a sample of institutions that was sparse may completely miss some asset types,
5) recovery rates by region, size of institution, and asset category were of equal importance and a close second to the national estimate in terms of how the data from the survey were going to be used.

Ordinarily, the most efficient procedure would be to set up an objective function (like a variance function) to minimize subject to a fixed cost. However, because so little was known about the relative costs of valuing the assets and the variation in the estimates of recovery by stratum, the most efficient procedure devolved to selecting an equal number of institutions per stratum, and an approximately equal number of assets within each sampled institution. In hindsight, because of problems in getting complete lists of assets from each sampled institution, this procedure was the correct one to choose. Any other procedure that would have involved more complicated sampling procedures at the asset sampling stage would have greatly slowed down the survey process.

For June of 1991, 60 institutions were sampled, about five per stratum (because of the initial distribution of institutions, one stratum had only four institutions sampled, and another had six sampled to compensate). These 60 were then assigned to four rotation groups, each group of size 15, to be recontacted in subsequent quarters. The 60 sample institutions were assigned so that all 12 strata were represented at least once in each quarter, and no stratum was represented more than two times. Within each institution, a list of all

assets in the 19 asset categories was obtained for sampling at the second stage, stratified by asset category. A sample of a minimum of five assets ultimately was selected in each asset category if there were five assets to be selected. If fewer than five assets were available in a category, all the assets in that category were taken into sample. Assets were ordered by book value (size) within a category and systematically sampled.

In September of 1991, the design became more complicated. There were now three sources or lists of institutions that were available. The first was the set of institutions that were originally contacted in June of 1991 that would now be recontacted in September. The same assets valued three months previously would now be revalued (because the status of the loan may have changed, the economic conditions affecting the sale price may have changed, or other factors may have had an impact in valuing the loan).

The second source was the set of institutions that were originally contacted in June of 1991 that would not be recontacted in September. Assets in this group of institutions would be "rolled forward", as describe earlier. Institutions in sources one and two represented all institutions resolved by June 1991.

For December 1991, we faced very much the same situation, except the first source of data was now the 15 institutions in the second rotation group from June, 1991 plus the 3 institutions assigned to rotation group 2 from September, 1991. From the second source of data, we roll forward the estimates from the remaining three rotation groups from June and September. The third source was again new sample from resolutions that occurred between August 1991 and November 1991, with assets sampled in the same was as in June of 1991.

March and June of 1992 proceeded in exactly the same fashion, but there were no new institutions added in March of 1992 because there were too few new resolutions. In June of 1992 we added 11 new institutions, again one per stratum. There were no institutions sampled in stratum four because there were no resolutions of institutions in that stratum.

Finally, for each administration of the survey, each asset was checked to determine whether it had been sold. This was done for rotation groups both in and out of sample so that information on sold assets could be obtained more rapidly and also to counter any biases that might occur for rotation groups if asset groups sold at differential rates.

Estimation

The final piece of the project is developing an estimation scheme. Standard estimation methods were used for most of the survey estimates, but some adaptations were made for determining confidence intervals.

For most of the recovery rate

estimates, we have quite a bit of information we can use to form ratio or regression estimators. Specifically, we know the book value of every asset, and for the longitudinal portions of the survey we know change in book value. At a minimum, we can use book value for all sampled and population assets, and the projected recovery for all sampled assets, and form the classical stratified ratio estimator for a two-stage clustered sample (Cochran, 1963). The ratio estimator and variance estimator for the ratio are both well-defined and have been known for a long time.

Because some of the recovery rates estimated are unusually low or high (close to zero or close to one respectively), it turns out that the tails of the confidence intervals produced as part of the survey estimation procedure go below zero (implying a negative recovery) or above unity (implying a recovery greater than the original value of the asset). While this may happen for an individual asset under very unusual circumstances, it cannot be true for the population value of the recovery rate for procedural reasons related to the methods used for selling the assets. This limitation means that the normal approximation so commonly used for construction of confidence intervals is appropriate for this survey.

As an alternative, we tried a Bayesian approach (Box and Tiao, 1973). We assumed that the recovery rate was a parameter drawn from a prior distribution, the Beta distribution. From the Beta distribution we determined the lower and upper bounds of the confidence interval so that we had the tightest confidence bound possible, with the restriction that the lower and upper bounds were in the interval on zero to one. This procedure in all cases reduced the confidence intervals over what would have been calculated using the normal distribution, but kept the mean and variance of the estimate the same.

Table 1: Asset Categories Used in Estimation of the Expected Cash Recovery

Assets Covered by the Survey
1) 1-4 Family Mortgages - Perf.
2) 1-4 Family Mortgages - Nonperf.
3) Multifamily (5+) Mortgages - Performing
4) Multifamily (5+) Mortgages - Nonperforming
5) Raw Land - Performing
6) Raw Land - Nonperforming
7) Construction Loan - Perf.
8) Construction Loan - Nonperf.
9) Commercial Mortgage - Perf.
10) Commercial Mortgage - Nonperf.
11) Commercial Loan - Perf.
12) Commercial Loan - Nonperf.
13) Consumer Loan - Performing
14) Consumer Loan - Nonperf.
15) Real Estate Owned
16) Furniture, Fixtures, and Equipment
17) Subsidiary Equity
18) Subsidiary Loans
19) Other Assets

Assets Not Covered by the Survey
20) Junk Bonds
21) Mortgage Backed Securities
22) Other Backed Securities
23) Judgements
24) Charge-Offs

Chart 1: Number of Institutions in ECR Sampling by Quarter

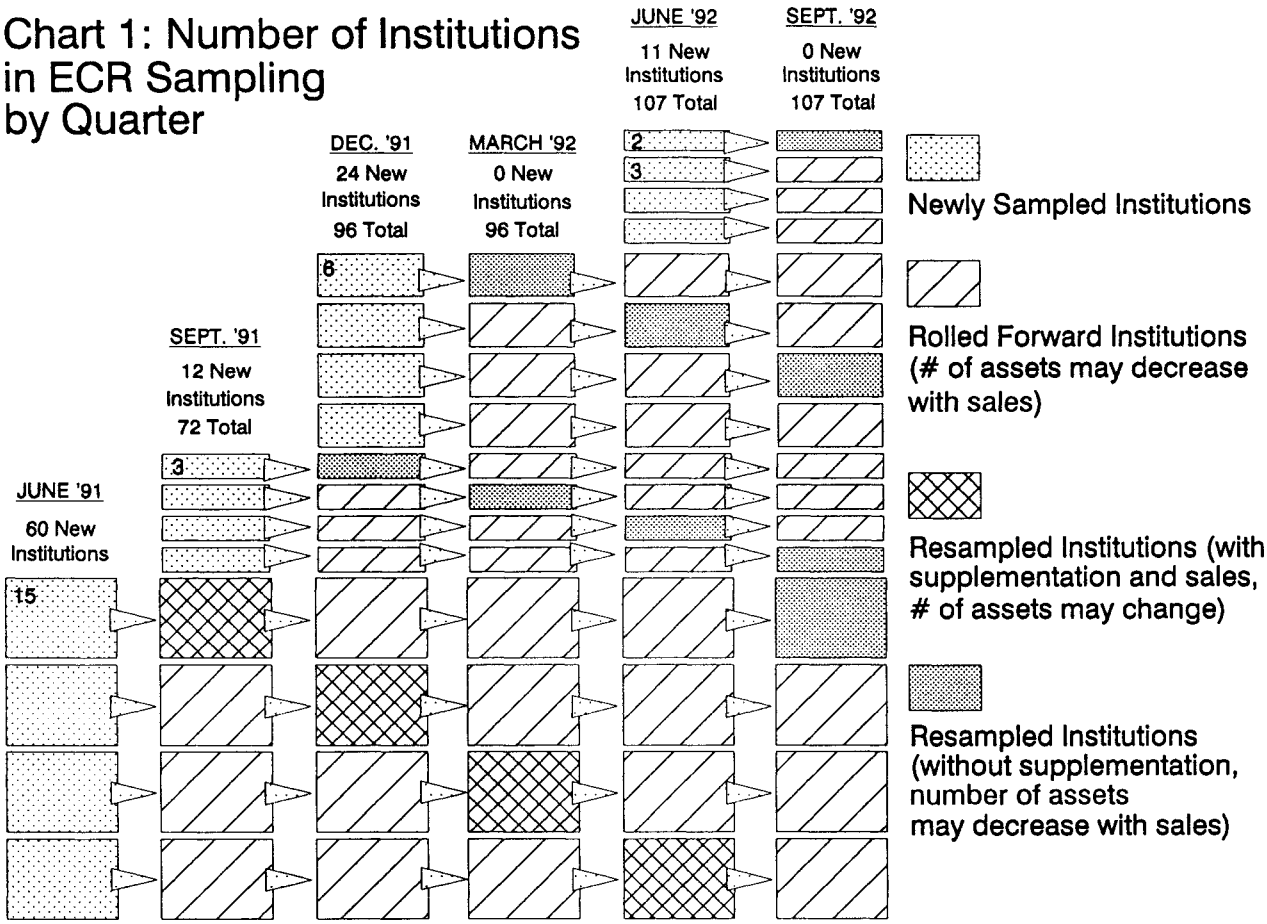

Chart 2: Regions for the Resolution Trust Corporation