EVALUATING REAL ESTATE VALUATION SYSTEMS

Robert J. Shiller, Yale University, Allan N. Weiss, Case Shiller Weiss, Inc.
Robert J. Shiller, Cowles Foundation, Yale University, Box 208281, New Haven, CT 06520–8281

Key Words: Real estate, property, valuation, appraisal, methods, testing, comparing, assessing, ranking, outcome, accuracy, error, mortgage, origination, strategy, default, foreclosure, Density Estimation and Profit Simulation (DEPS) method

For a real estate valuation system (or appraisal method) to perform well, it should help the mortgage originator, who uses the system to judge whether to approve mortgage applications, prevent default-related expenses and loss of principal without incurring too many other costs, and put the lender at an advantage, appropriately defined, in terms of other firms. Unfortunately, while this objective is simple to state, there are many difficulties in translating it into a criterion for judging real estate valuation systems. There are many complicated links from the real estate valuation systems to this ultimate costs and benefits to the mortgage lender. Because of these difficulties, faulty decisions are being made, in making comparisons between competing statistical valuation systems or between a purely statistical valuation system and an appraisal.

At this time of rapid development of automated underwriting systems, the importance of evaluating real estate valuation systems is greater than ever, and at the same time, because the competitive situation for mortgage lenders is very complex, and automated underwriting systems are themselves in flux, the costs and benefits to competing valuation systems are difficult to define.

We assume here that data are available for testing valuation systems; the data are assumed to be in the form of current estimated values rendered by each of the competing valuation systems (including appraisal methods) for individual properties, actual unforced (not defaulted) selling prices, current as well as subsequent, for the same properties, loan amount, as well as other information that may be available about the properties or the valuations. Data sets with some of this information are often constructed by mortgage originators when they are considering changing their valuation systems.

Many tests evaluating valuation systems using such data do nothing more than compare measures of dispersion of the errors, of the percentage difference of an actual sales price relative to the predicted price. While these tests are useful, there are many reasons why they may be misleading, and other measures of valuation system success are also important.

We provide here a framework that will enable us to compare different valuation systems (including comparing appraisals with statistical valuation systems) in a manner that is directly revealing of the results for the mortgage underwriting process. The framework that we propose has two aspects. First, we use the framework to develop a Density Estimation and Profit Simulation (DEPS) Method that, given an assumed relation of mortgage lender profits to the true loan-to-value ratio, provides a direct estimate of the expected value of the valuation system to the mortgage lender.Second, the framework suggests various simple measures of the quality of valuation systems that are generally overlooked today.

1. Type I and Type II Error

Our analysis of the usefulness of real estate valuation systems must begin with the recognition that there are two kinds of collateral valuation errors that are costly to mortgage lenders. Borrowing terminology from theoretical statistics, we use the term type I error to refer to the error of rejecting a mortgage applicant who would not have defaulted. The costs to this error are the cost of the loss of profit from writing a good mortgage, and the loss of good will and further business and referrals from the rejected applicant, as well as the cost of the underwriting process (including the cost of purchasing the real estate valuation) when no loan is written. We use the term type II error to refer to the error of accepting a mortgage applicant who defaults. The costs to this error are the costs of problem-loan servicing, including collection efforts, cash flow advances, foreclosure administration, legal costs, and principal losses related to a lender's sales of the foreclosed property (as well as the cost of the original underwriting process).

Presumably the potential cost for each property from type I error is less than that of type II error: potential default loss is greater than poten-

Reprinted from the 1996 Proceedings of the Business and Economic Statistics Section of the American Statistical Association

tial profit from writing a good mortgage. Thus, in some sense one might think that type I error is less important than type II error. However, from the standpoint of comparing valuation systems the type I error is no less important since the approval criteria are set so that type I error will occur more often, at the margin, than does type II error. We must consider both kinds of error in evaluating approval methods. If we considered only type I error, we could reduce this to zero by selecting a valuation system that always gives a very high value for each property, so that no applications are ever rejected. If we considered only type II error, we could reduce this error to zero by selecting a valuation system that always gives a very low value for each property, so that all applications are rejected. Neither of these extremes makes any sense for mortgage lenders.

Simple variances of real estate valuation errors are not sufficient statistics for the type I and type II errors that are relevant to judging the valuation systems. They do not take account of the shape of the distribution of errors. Taking full account of the distribution of errors of the valuation system is not something that can be done formally until we specify the precise use of the valuation in the underwriting method, and the nature of the costs associated with underwriting errors.

2. The Simplest Model for the DEPS Method

Let us consider first a very simple profit simulation model that allows us to take account of type I error and type II error with the DEPS method, and that suggests some of the complexity of the problem of comparing valuation systems. The model assumes that the mortgage lender operates in isolation, so that there is no issue of selection bias that would arise if the lender's underwriting guidelines permit acceptance of borrower loan applications that other lenders reject. This simple profit simulation model also makes no use of the selling price of the property itself: we will assume that we are either applying our analysis to refinance mortgages, for which no current purchase price is available, or our mortgage applications have already been prescreened on the basis of selling price relative to loan applied for.

The value v is the natural log of the true value of the property (price it would obtain in a normal sale, not a foreclosure sale) scaled by subtracting the natural log of the loan balance applied for. We will use the symbol $\hat{v}$ to refer to the natural log estimated value of the property also scaled by subtracting the natural log of the loan balance applied for. Let $f(v,\hat{v})$ be the joint probability density function for v and $\hat{v}$ among the loan applications that the mortgage lender wishes to consider with a valuation, including both applications that will be accepted and applications that will be rejected on the basis of the valuation. A simple measure of the quality of the valuation system, similar to measures that are widely used to evaluate valuation systems today, is σ_1^2, the second moment, or if $\hat{v}$ is unbiased, the variance, of $v - \hat{v}$. In terms of $f(v,\hat{v})$ this is:

$$\sigma_1^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(v-\hat{v})^2 f(v,\hat{v})\,dv\,d\hat{v} \qquad (1)$$

The valuation system with the lowest σ_1^2 is judged the best. This simple measure will be contrasted with other measures.

Let us use the symbol π to represent the expected profit as a fraction of the loan amount applied for with a single mortgage application. (π is not in logs, as are v and $\hat{v}$.) To allow us to account for the relevant costs of the two types of error, let us suppose that π is a function of v. We suppose that the function $\pi(v)$ is concave down, is very negative for small values of v (reflecting such factors as foreclosure costs and the loss of property value in a foreclosure sale), but is only slightly positive for large values of v, and becomes flat (π becomes unaffected by v) for high values of v. This shape for the function embodies our assumption that costs of type II error (of accepting an application that was bad because of low v) can be much larger than the benefits of accepting a good application.

The mortgage lender decides on a threshold value $\hat{v}^*$ so that only mortgage applications for which $\hat{v}$ exceeds $\hat{v}^*$ will be accepted. The expected profit Π that the mortgage lender obtains, accounting for the profit function, the joint distribution $f(v,\hat{v})$ using this valuation as well as the cost C of obtaining the valuation $\hat{v}$ (including associated application processing costs) is then:

$$\Pi = \int_{\hat{v}^*}^{\infty}\int_{-\infty}^{\infty}\pi(v) f(v,\hat{v})\,dv\,d\hat{v} - C \qquad (2)$$

The valuation system with the highest Π is judged the best. Note that the expected profit Π is not the expectation of $\pi(v)$ conditional on $\hat{v}$ being greater than $\hat{v}^*$, since such a conditional expectation would not take account of the probability of the type I error of rejecting a good mortgage. The expression embodies the assumption that the cost C of purchasing the valuation is borne even if the mortgage application is rejected. The expected profit

Π is a sort of weighted integral of the probability density function, $f(v, \hat{v})$, with weights $\pi(v)$ that are very different from the $(\hat{v} - v)^2$ that appeared in equation (1). The isoquants of $(v - \hat{v})^2$ are (plotting v on the horizontal axis and $\hat{v}$ on the vertical axis) parallel straight lines with slopes of one, while the isoquants of $\pi(v, \hat{v})$ are parallel straight lines with slopes of infinity; only the region above of $\hat{v}^*$ is used. This shows immediately that the measure of the quality of the valuation system may be very different between (1) and (2).

It must be recognized that $\hat{v}^*$ is a choice variable for the mortgage lender. Assuming that the mortgage lender operates so as to maximize expected profits, then the lender will choose $\hat{v}^*$ to maximize (2), i.e., so that $\partial\Pi/\partial\hat{v}^* = 0$, equation (3),

$$\int_{-\infty}^{\infty} \pi(v) f(v, \hat{v}^*) dv = 0 \qquad (3)$$

and the mortgage lender should proceed to obtain the valuation only if the optimized Π is positive.

Loan applications where $v = \hat{v}^*$ are the marginal loans, for which the expected profits earned when the decision to lend proves successful are exactly matched by expected losses when the decision to lend proves detrimental. The lender is indifferent between making these loans or not making them, but for mortgage applications where $v > \hat{v}^*$ the mortgage lender can expect to make a profit. Note that the choice of $\hat{v}^*$, made as it is with expression (3) which involves $f(v, \hat{v})$, depends on characteristics of the valuation system, and so $\hat{v}^*$ will differ across valuation systems. This variation across valuation systems in the optimal $\hat{v}^*$ is very important to account for: one valuation system may have advantages over another only because it allows the mortgage lender to lower $\hat{v}^*$ and thereby lower type I error.

Given this model, we propose that a good method to compare valuation systems, the DEPS method is, for each valuation system, to estimate the joint distribution $f(v, \hat{v})$ using data on actual values of properties and their estimated values using the valuation system, and loan balances applied for, to hypothesize a profit function $\pi(v)$, to use equation (3) to derive the $\hat{v}^*$ for that valuation system, and then use equation (2) to derive Π. The valuation system with the highest Π would be, in terms of this simple model, the best method. We suppose that a nonparametric method or a very general parametric method of estimating the probability density function should be used if there are sufficient data, so that the shape of the function can be captured in some detail; see Tapia and Thompson [1978], Révész [1984] or Silverman [1986].

Consideration of this model also suggests simple measures that can be obtained without going through the DEPS method, measures beyond just variance of valuation errors, that would be considered as reflecting on the quality of a real estate valuation system. One such measure, which might be looked at in addition to the variance shown in equation (1), is the skewness of the distribution of $v - \hat{v}$. Suppose we are comparing two real estate valuation systems both which are unbiased ($E(v - \hat{v})$ equals zero) and both which have the same variance given by equation (1). We would then generally prefer the valuation system whose error $v - \hat{v}$ is skewed to the left, i.e., whose large errors tend to be to undervalue a property. Really large undervaluation errors tend not to hurt the mortgage lender, since if the errors were somewhat smaller then the mortgage application would be rejected anyway, and so there is no difference in costs to the mortgage lender. On the other hand, really large positive errors can be catastrophic to mortgage lenders if they cause lenders to issue mortgages on properties whose values are much less than their respective loan amounts. That is why negative outliers matter less than positive outliers. Thus, valuation systems with negatively skewed distributions of $v - \hat{v}$ tend, other things equal, to perform better in the DEPS method with this model.

Measuring the skewness of the distribution of $v - \hat{v}$ is not a substitute for our DEPS method. For example, skewness might be related to the level of $\hat{v}$, and only when $\hat{v}$ is above $\hat{v}^*$ does skewness matter. One could derive other revealing simple measures, more closely related to the DEPS method, of the quality of a valuation system by simplifying the π function so that it is, for example, a step function. We might suppose that for v below some threshold v^- π is the negative constant π^-, and above some threshold v^+ (where v^+ is greater than v^-) π is the positive constant π^+, and π is zero between these two values. Then $\Pi = \pi^- \text{Prob}(v < v^- \text{ and } \hat{v} > \hat{v}^*) + \pi^+ \text{Prob}(v > v^+ \text{ and } \hat{v} > \hat{v}^*) - C$. Then, useful summary statistics would be the two estimated probabilities. Users could use their own values for π^-, π^+, and C to evaluate Π given these probabilities.

Reporting such measures, the skewness or the probabilities defined just above, is especially important in comparing statistical valuation systems with appraisals. Using measures of dispersion of valuation errors alone to make comparisons may be unfair to the appraisers, since appraisers operate with a sense of the costs of type I and type II error, and tend to be (correctly) mindful of these costs in allocating their time and effort in the appraisal process. In appraising properties where the value is obviously sufficiently above the loan amount, the appraiser may see less benefit in expending extra effort to refine the valuation. The result could be a misleadingly high standard deviation of errors for appraisers.

3. Correlation of Valuation Error with Selling Price Error

In the preceding section, we disregarded the fact that for purchase mortgages, in contrast to refinance mortgages, the mortgage lender actually has two pieces of information: the valuation placed on the property, which provides $\hat{v}$, and the selling price today. We will use the symbol p to represent the natural log of the selling price today scaled by subtracting the natural log of the loan balance applied for. Note that the true log value used to compute v is not the current selling price, but is the price at next sale, on a date a few years down the road when mortgagor default propensity is greatest.

Let us define $f(v, \hat{v}, p)$ as the joint probability density function of the three quantities among all purchase mortgage applications that the mortgage lender is considering, even including those applications that might be ultimately rejected because the loan amount applied for is too high relative to the current selling price. The most common measure of the accuracy of valuation systems used in practice is actually σ_2^2, the second moment, or, if $\hat{v}$ is unbiased, the variance, of $p - \hat{v}$. This is not the variance σ_1^2 of $v - \hat{v}$. It is instructive, for comparison, to write σ_2^2 in terms of $f(v, \hat{v}, p)$:

$$\sigma_2^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(p - \hat{v})^2 f(v, \hat{v}, p)dv\, d\hat{v}\, dp \quad (4)$$

As before, by the conventional criterion, the valuation system with the lowest σ_2^2 is judged the best.

In contrast, the expected profit to a mortgage lender that accepts a loan application for which p is above some threshold value p^* and for which $\hat{v}$ is above $\hat{v}^*$ is:

$$\Pi = \int_{p^*}^{\infty}\int_{\hat{v}^*}^{\infty}\int_{-\infty}^{\infty}\pi(v) f(v, \hat{v}, p)dv\, d\hat{v}\, dp - C \quad (5)$$

and, as before, the valuation system with the highest Π is judged the best. Note that, much as in the preceding section, the expected profit Π is a sort of weighted integral of the probability density function, here $f(v, \hat{v}, p)$, with weights, determined by $\pi(v)$ and the thresholds $\hat{v}^*$ and p^*, that are totally different from the $(p - \hat{v})^2$ that appeared in equation (4).

As with the simpler model of the preceding method, the thresholds $\hat{v}^*$ and p^* are both decision variables for the mortgage lender. Taking the partial derivatives of the expected profit, equation (5), with respect to $\hat{v}^*$ and p^* and setting these to zero, we have the pair of equations:

$$\int_{\hat{v}^*}^{\infty}\int_{-\infty}^{\infty}\pi(v) f(v, \hat{v}, p^*)dv\, d\hat{v} = 0 \quad (6)$$

$$\int_{p^*}^{\infty}\int_{-\infty}^{\infty}\pi(v) f(v, \hat{v}^*, p)dv\, dp = 0 \quad (7)$$

These two equations, solved together, will yield the optimal thresholds $\hat{v}^*$ and p^* for purchase mortgages. Note that changing the valuation system will in general cause changes in both $\hat{v}^*$ and p^*.

Proceeding as in the preceding section, the DEPS method of comparing valuation systems is, for each valuation system, first to estimate the probability density function $f(v, \hat{v}, p)$, then to use equations (6) and (7) to derive the optimal $\hat{v}^*$ and p^*, and finally to evaluate (5). The valuation system with the highest Π will be judged best. The model we have defined is actually rather complex: in comparing just two valuation systems there are four thresholds to be derived, and getting these wrong in the comparison process could reverse the conclusions.

A simpler statistic that bears on the results that we will obtain using the DEPS method with this model is the correlation of the valuation error $v - \hat{v}$ with the current selling price error $v - p$. It is very important to realize that errors in valuations usually do not result in default losses among purchase mortgages: the losses tend to occur only when the selling price of the house is too high relative to true value *and* the valuation is also too high relative to true value; only then will the mortgage lender be influenced to make a costly error. Selling price error is the error made by the buyer of the house, paying too much for the house relative to its true market value, that is, too much relative to what the typical buyer would pay. If this error is independent of the valuation system error, then the probability of a loss is related essentially to the product of the probabilities of the two large errors; if both probabilities are small, this product will be much smaller. It is important to note this, since a tendency to make occasional large errors is common among statistical valuation systems, and yet this tendency is really not very costly if the errors are uncorrelated with current selling price errors.

By assuming here that true value v is represented in empirical work by the unforced sale price a few years after the mortgage is issued, this version of the DEPS method builds into the evaluation method an incentive for those who provide valuations to provide not estimates of current value of properties but instead forecasts of future values. Many appraisers do not describe what they do as forecasting, but there is a recognition in the appraisal industry that forecasting of some sort is inevitably involved, see Lovell [1994] or Shlaes [1992].

4. Models Involving Games Against Competing Firms

In applying a valuation system, one must recognize that one's method is not used in isolation: other mortgage lenders are also using valuation systems, and the benefit to a given valuation system will tend to depend on the nature of the valuation system used by competitors. The risk that a mortgage lender faces is that it will fail to see some information that caused other mortgage lenders correctly to deny mortgage credit. It might become a haven for all such rejected applicants. Alternatively, if the mortgage lender can find a way to use the information in the valuation system to find good applications among those rejected, there may be a very large profit opportunity.

A valuation system that underperforms the method used by competitors in terms of variance of errors might, by the DEPS method, be better than a valuation system that outperforms the competitors in terms of the variance of errors. This can occur if the valuation system is in effect using different information than that of the competitors. Obviously, a valuation system that gives the same valuations as the competitors' will produce zero profits for the mortgage lender by this model. At another extreme, if the information used by the valuation system is completely different from that used by competitors but no better in terms of variances of forecast errors, then there could be large expected profits for the mortgage lender. This suggests that a simple measure of the value of the valuation system would be the correlation of the error made by the valuation system with the error made by competitors' valuation system, the smaller the better.

5. Other Considerations

The DEPS method can be modified for analysis of each of the following considerations:

1. Risk Considerations For a risk-neutral mortgage loan underwriter, the correlation between valuation errors of different properties represented in its portfolio of mortgages is of no account, because the expected losses are the same whatever the correlation. However, if errors are correlated across properties, then the variance of the underwriting losses to the portfolio can be magnified, and this magnification can matter to risk-averse underwriters. The more a valuation system uses outdated information, the more, for a given variance of errors, the valuation errors are likely to be correlated across properties.

2. Costs of Failure to Make a Valuation Many valuation systems are often unable to make a valuation at all, since the information on which the method is based may not be available. Alternatively, the producer of the valuation may be computing a standard error for the valuation, and only reporting those valuations that have a standard error below some threshold decided on by the valuer, so that the errors on reported properties will tend to be small.

3. Making Use of the Standard Error of the Valuation Some valuation systems may provide estimated standard errors of the valuations, which inform the mortgage lender how good the information is about the true value. These standard errors, if they vary from property to property as a reflection of the differing quality of information about them, are potentially very useful information to the mortgage lender, since the values $\hat{v}^*$ and p^* can be made a function of them.

4. Using Multiple Valuation Systems In the case where a valuation system is found to contain information that is not contained in other valuation systems, it should be explored whether there are profits to be gained by using both methods.

6. Conclusion

We have seen that conventional measures of the success of valuation systems, such as measures of dispersion of valuation errors, often are unfair tests of the usefulness of the valuations to the mortgage lenders. Using such measures for comparisons of statistical valuation systems with conventional appraisals are especially likely to be unfair, since appraisers, in providing a service to mortgage lenders, have in mind the relative costs and likelihood of the different kinds of errors, not taken into account in the conventional measures. The DEPS method proposed here will correct these deficiencies of the conventional measures of success.

REFERENCES

Case, Karl E., and Robert J. Shiller, "The Efficiency of the Market for Single Family Homes," *American Economic Review*, 79:125–37, 1989.

Case, Karl E., Robert J. Shiller, and Allan N. Weiss, "Mortgage Default Risk and Real Estate Prices: The Use of Indexed-Based Futures and Options in Real Estate," National Bureau of Economic Research Working Paper, (*Journal of Housing Research*, forthcoming), 1995.

Isakson, Hans R., "The Nearest Neighbors Appraisal Technique: An Alternative to the Adjustment Grid Method," *Journal of the American Real Estate and Urban Economics Association*, 14(2):274–86. 1986.

Ito, Takatoshi, and Keiko Nosse Hirono, "Efficiency of the Tokyo Housing Market," *Bank of Japan Monetary and Economic Studies*, 11:1–32, 1993.

Lovell, Douglas D., "Improving Your Appraisals by Avoiding Forecast Errors," *The Appraisal Journal*, 60(3):424–6, July, 1992.

Poterba, James, "House Price Dynamics: The Role of Tax Policy and Demography," *Brookings Papers on Economic Activity*, 2:143–99, 1991.

Révész, P., "Density Estimation," in P. R. Krishniah and P. K. Sen, eds., *Handbook of Statistics*, North Holland, Amsterdam, 1984.

Shlaes, Jared, "Real-Time Appraising: Bringing Value Up to Date," *The Appraisal Journal*, 60(3):419–24, July 1992.

Silverman, B. W., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, 1986.

Tapia, R. A., and J. R. Thompson, *Nonparametric Density Estimation*, Johns Hopkins University Press, Baltimore, 1978.