Lenders are assumed to use formal credit scoring schemes in order to evaluate borrower credit worthiness. Variables used in these schemes may be measured with error resulting in credit scores which include the effects of biased parameter estimates, and in lending decisions that appear to be discriminatory although lenders are not prejudiced. Regulations which restrict the information used in credit scoring schemes may produce undesirable credit supply results. Theoretical models are supplemented with illustrative empirical analysis of mortgage lending in which use of information on property location is prohibited. The empirical results indicate that the quantitative impact of such regulations is modest.

I. INTRODUCTION

In recent years federal and state legislatures increasingly have become interested in the criteria which lenders use to decide upon the amount of credit to extend to individual borrowers. In response to borrower allegations of arbitrary or discriminatory treatment when applying for loans, major regulations have recently been enacted to limit the type of information lenders may use to determine which borrowers are acceptable credit risks. The Equal Credit Opportunity Act of 1976 and the Community Reinvestment Act of 1977, for example, either expressly prohibit or specifically discourage the use of certain information to determine credit-worthiness. In principle, these laws are intended to remove non-economic barriers to credit for various groups which would otherwise qualify for credit. However, in some non-trivial cases such regulations may produce the opposite effect. Chandler and Ewert (1975) and Chandler and Coffman (1980), for instance, have argued that women on average actually are better credit risks than men. Forcing creditors to ignore sex in assessing credit-worthiness may therefore, reduce credit supply to the very group that supposedly is being protected.

This paper examines analytically and empirically the potential effects of regulations that restrict the information set available to financial institutions when making lending decisions. The next section of the paper presents a simple model of the credit market which demonstrates the role of credit-scoring schemes and econometric default loss models in lender decisions. Section III examines explicitly the effects of limiting the use of certain types of information on a lender's determination of the credit-worthiness of a borrower, and hence, on the overall impact of such limitations on lender behavior. In section IV the potential effects of informational restrictions are quantified using unique estimates of a default loss equation for mortgage loans. The final section summarizes the principal conclusions and suggests some policy implications of our analysis.

*The George Washington University. The authors are Professor of Economics and Associate Professor of Economics at the George Washington University. Helpful comments on earlier versions of the manuscript were provided by Warne Rutland, Robert Ellickson, Anthony Santomero, and an anonymous referee. The data research assistance of John Duncan and Andrew Parks is gratefully acknowledged. This research was financed by National Science Foundation grant DAB 078-0887. The views expressed here are those of the authors.

Economic Inquiry
Vol. XXI, July 1983

349
The role of credit serving in the lending decisions made by financial institutions may be illustrated with a simple model of the consumer loan market. The lender is assumed to be a multiproduct firm in an industry without any barriers to entry. Products provided by the lender include loans which differ in amount, interest rate, collateral, term, and default recourse. The expected profit to a lender for a given individual loan transaction may be expressed as:

\[ E(P) = R(i, L, T, r) - E(D(L, i; C, T, X, Y)) - [I(L; C, X, Y)], \]

where \( E(P) \) is the expected profit on an individual loan transaction, \( R(i, L, T, r) \) is the revenue, net of capital cost, on a loan of \( L \) dollars for \( T \) years, at interest rate \( i \), with capital cost \( r; E(D(L, i; C, T, X, Y)) \) is the expected default loss on a loan of \( L \) dollars for \( T \) years, at interest rate \( i \), with collateral \( C \), with other terms \( X \), and with borrower characteristics \( Y \); and \( I(L; C, X, Y) \) is the information and processing cost on a loan of \( L \) dollars, with collateral \( C \), with other terms \( X \), based on the use of information represented by borrower characteristics \( Y \).

Equation (1) indicates there are many factors to be evaluated in negotiating a loan. However, for expository purposes, it will be assumed that loan products with the same term to maturity (\( T \)), collateral (\( C \)), and recourse (\( X \)) provisions can be grouped together as the same product. A mortgage loan with some specific real property as collateral would be a different loan product than an unsecured personal loan. With \( T, C, \) and \( X \) held constant for each particular loan type, the lender determines the optimal values for \( L, i, \) and \( Y \).

If lenders are unprejudiced and risk-neutral, the supply of loanable funds of different types (i.e., with different combinations of \( T, C, \) and \( X \)) would be based on maximization of \( E(P) \) with respect to \( L, i, \) and \( Y \). Moreover, in the absence of barriers to entry, lenders would earn only normal profits, which are included in \( r \), so that \( E(P) \) would be driven to zero. Based upon these considerations the supply schedule of a particular loan product, holding \( T, C, \) and \( X \) constant, would consist of combinations of \( L, i, \) and \( Y \) that produce a constant (zero) value for \( E(P) \). By varying \( T, C, \) and \( X \) one obtains the appropriate supply schedule for different loan types. If lenders fail to offer all possible loan products, some transactions would not take place because lenders consider certain product lines to be unprofitable even though these may be desired by some borrowers.

For any given loan product, the relevant supply schedules for different borrowers may be truncated, due to legal regulations such as usury ceilings, or non-price rationing caused by imperfect information, and/or borrower self-selection. This means that the amount of credit for which any given borrower qualifies will be determined by the relationship between the range of interest rates offered by the lender and the (zero-profit) supply schedule appropriate for any particular borrower.

1. This assumption which simplifies the analysis of the varying terms of consumer loans has been used in Bitter and Dejonghe (1973).
2. This is a common assumption. See Jaffee and Russell (1976) and Barth, Corlet, and Yuse (1979).
3. See, for example, Stiglitz and Weiss (1981).
Figure 1 depicts the case of a lender supplying a particular type of loan at interest rates not exceeding a legal rate ceiling of \( k \). The curves \( S_1 \), \( S_2 \), and \( S_3 \) represent zero-profit supply schedules associated with loans offered to three applicants, A, B, and C, respectively.\(^4\) Clearly, A would not qualify for any credit, while B and C would qualify, though at different terms, with B paying a higher interest rate and/or receiving less credit than C.

**FIGURE 1**

Individual Credit Supply Given a Binding Interest Rate Ceiling

Differences in the magnitude of the expected default loss, \( E(D(L, i, T, X, C, Y)) \), may cause the position of the supply schedules \( S_1 \), \( S_2 \), and \( S_3 \) to differ. For given values of \( L, i, T, X, \) and \( C \) these differences would reflect the personal characteristics of the three borrowers, \( Y_1, Y_2, \) and \( Y_3 \). Knowledge of these characteristics could be part of the lender’s information set. More specifically, in determining the optimal information set, the marginal cost of acquiring additional information would be balanced against the marginal benefits. The latter would be defined in terms of either higher expected profits, or if the lender were risk-averse, a reduction in the variance of \( E(P) \). In our analysis, we assume that this optimal information set is chosen in a manner consistent with the formal criteria derived in Kihstrom (1974), Aigner and Sprenkle (1968), and Stanhouse and Sherman (1979).

Once the optimal information set, \( Y^* \), is determined the lender would transform the borrower characteristics into a prediction of expected default loss. For example, let \( d \) be a random variable representing the default loss on a loan and \( Y^* \) be a vector of

\(^4\) A significant increase in \( d \) with increasing \( i \) can be expected, for secured loans, at the point where the principal approaches the value of the collateral. The explicit model of borrower behavior implied by the shape of the supply schedule is found in Jaffe and Russell (1978) and Barth, Contie, and Tiner (1979). See Benston (1977) for an analysis of the effects of risk on loan supply.
borrower characteristics reflecting the lender's perception of the optimal information set. The lender attempts to obtain the best prediction of \( d \) conditional on the value of \( Y_0 \). However, Spence (1974) has shown that, under certain assumptions, a credit score, \( S \), formulated as a weighted sum of the \( Y_0 \)-variables, \( S = \sum b_i (Y_i - \bar{Y}_i) \), would be an efficient predictor of \( d \) given \( Y_0 \). That is, the conditional distribution of \( d \) given \( S \) is identical to the conditional distribution of \( Y_0 \). The lender could use the credit score of an applicant as a predictor of the default loss appearing in equation (1). Lending decisions, concerning the interest rate and/or the rejection of the applicant, would then be made so as to produce ordered pairs of \((i, E(d))\) or \((i, S)\) that yield expected profits in the range determined by the competitive conditions in the industry.

III. EFFECTS OF INFORMATIONAL REGULATION ON CREDIT SCORING AND CREDIT SUPPLY

In the analysis which follows, we assume that the amount of credit supplied to an individual borrower is based on a credit score computed from an estimated default loss equation. We further assume that lenders do not discriminate against certain types of borrowers on the basis of non-economic factors. Though lenders are assumed to be unprejudiced, they may be either risk-neutral or risk-averse. We then compare the behavior of lenders when no limitations are placed on the information which may be used, but key informational variables are measured with error, with the behavior resulting when lenders are prohibited from using certain information.

Consider first the case in which some of the variables required to estimate the default loss equation to be used in the credit scoring scheme are measured with error. An example of such a borrower characteristic is the monthly payment to income ratio (hereafter MPRI). Expected default will clearly depend on future values of MPRI, but the lender will typically only be able to determine the current value of MPRI. For simplicity, assume that the lender's optimal information set contains the single observed borrower characteristic, \( Y_i \), or, \( MPRI \), that \( Y_i \) may be used freely in both default loss and credit scoring models; and that the true default loss model has the form:

\[
d = B_0 + B_1 y_i + u,
\]

where \( d \) is the default loss in dollars; \( y_i \) is the true value of observed borrower characteristic \( Y_i \); and \( u \) is a normally distributed random error term uncorrelated with both \( y_i \) and \( Y_i \). Assume further that the lender regards any individual loan applicant, \( i \), as a random selection from the population of previous borrowers for whom data on defaults and borrower characteristic \( Y_i \) are readily available.

If the observed and true values of this borrower characteristic differ by a normally distributed random error, \( \epsilon_i \), so that:

\[
Y_i = y_i + \epsilon_i,
\]

the default loss equation estimated from the population of previous loans may be written in terms of parameters of the true model as:

\[
d = B_0 + B_1 Y_i + (u - B_1 \epsilon_i).
\]
A well-known result is then:

\[ \text{plim} \, b_i = \frac{\text{cov}(d, y_i)}{\text{Var}(y_i)} = B_i / (1 + (\text{cov}(y_i, \text{cov}(y_i)))) \]

where \( b_i \) is the estimate of \( B_i \).

To constructing the credit score, the lender is interested in the conditional expectation of \( d \) given \( Y_i \). If \( d \) represents the predicted value of default loss from equation (4) then:

\[ d_i = d + b_i(Y_i - \bar{Y}) \]

where \( d \) and \( \bar{Y} \) are group means and the subscript \( i \) indicates the borrower being evaluated. Since \( \text{plm} \, d = E(d|Y_i) \), prediction from the least squares regression yields consistent estimates of the conditional expectation of the default loss given \( Y_i \).

Although the use of equation (6) as a credit scoring scheme would yield correct results on average, the supply of credit would not be the same as when the true value of \( y_i \) was observed and a consistent estimate of \( B_i \) obtained. To some advocates of regulation, this divergence might appear to be discriminatory. Since plim \( b_i/B_i < 1 \), the default losses of high \( MPAY \) borrowers \( (Y_i > \bar{Y}) \) will be underestimated, and those of low \( MPAY \) borrowers \( (Y_i < \bar{Y}) \) will be overestimated. Such a situation would appear to imply a bias against (low) borrowers with a low (high) \( MPAY \) for whom default is overestimated (underestimated). This appearance would, however, be misleading since the differences between \( Y_i \) and \( y_i \) is more likely to be positive (negative) for the borrower with a high (low) \( MPAY \) and the lender would quite properly wish to take this into account.

Suppose that the use of \( Y_i \) to predict default were prohibited because the outcome appeared to be unfair or discriminatory. If alternative \( Y_i \) were not available, the lender would then base credit supply on a default model and a credit scoring scheme in which \( d = d \) for all borrowers. If the lender were risk neutral, this would not affect the supply of credit to the average borrower. However, interest rates would fall (rise) for borrowers with a high (low) \( MPAY \) above average (below average). If the lender were risk averse, credit supply would vary inversely with \( E(P) \), which would increase if the lender were prohibited from using a statistically significant variable in predicting default. The regulation of a risk-averse lender would therefore raise the general level of interest rates faced by the average borrower in addition to altering the distribution of interest rates charged different borrowers.

Actual informational regulations have concentrated on the use of variables such as property location, sex, and race which are not likely to be measured with error. Consider, for example, \( X \) a binary characteristic, \( Y \), dividing applicants into two groups, \( A \) and \( B \), so that the optimal information set consists of \( Y \) and \( Y \). Assume

5. See Johnston (1982).


7. The model presented here may also be extended to the case in which \( Y \) becomes a vector. If only one variable in the \( Y \) vector is measured with error, the results pertaining to the direction of bias are unchanged. However, when more than one variable is measured with error, the determination of the direction of bias becomes very complex and the effect of having a particular variable from the \( Y \) vector become as empirical issue.
further that unlike $Y_1$, measured $Y_1'$ differs from true $y_1$ by a random normal measurement error $\epsilon_1'$ (and similarly for $B$).

In this case, the pair of default loss equations used to create the credit scores would be:

$$(7) \quad d' = B_1' + b_1'Y_1' + (u' - B_1('e_1')),$$

and

$$(8) \quad d' = B_2' + b_2'Y_2' + (u' - B_2('e_2')).$$

The corresponding credit scores would be:

$$(9) \quad S_1' = \tilde{a}_1' - \tilde{d}' + b_1'(Y_1' - \bar{Y}) = E(d_1'|Y, Y_1);$$

and

$$(10) \quad S_2' = \tilde{a}_2' - \tilde{d}' + b_2'(Y_2' - \bar{Y}) = E(d_2'|Y, Y_2).$$

Since $plim b_1' = E/(1/(1 + (var(c_1')/var(y_1')))$ and similarly for $plim b_2'$, both estimates would be biased and inconsistent.

Consider first a case in which the lender’s use of information is not regulated, and groups $A$ and $B$ differ only with respect to the mean default loss with $\bar{d}_1' > \bar{d}_2'$. As shown in figure 2, separating applicants from group $A$ and $B$ in terms of characteristic $Y$, would be tantamount to assigning an applicant to a point on either regression line $A$ or $B$.

**FIGURE 2**

Restricting the Use of Information About $Y_1$
As Phelps (1972) and Aigner and Cain (1977) have shown in models of labor markets, this would result in different treatment for applicants with identical Y, based on membership in either A or B. Some observers would regard this as discrimination, though such differential treatment is fully consistent with unprejudiced, profit-maximizing behavior.

If the use of Y were banned, lenders would be compelled to estimate a single, pooled default loss equation, AB. Risk-neutral lenders would charge the same interest rate to both A and B borrowers with equal Y values, thereby raising interest rates charged group B applicants and lowering rates charged group A applicants. Risk-averse lenders would respond to an increase in variance from var(d|Y1, Y2) to var(d|Y1), by shifting their credit supply schedules upward.

Somewhat different results obtain if we assume that the only difference between groups A and B is that the measurement error in observing Y is greater for group A, (i.e., var(e)) > var(e)). As shown in figure 3, an unregulated lender would continue to estimate separate default loss equations, and credit score equations; but in equations (8) and (10) \( d^* = d^A \) and \( \text{plm} b^* < \text{plm} b^A \).

**FIGURE 3**
Restricting Use of Y, When Y, Measurement Error Diffs

When \( Y_1 = Y_1 > Y_2 \), members of group A obtain more favorable credit terms from a risk-neutral lender because their expected default loss is lower. This ranking is reversed for \( Y_2 = Y_2 < Y_1 \). By comparison, a risk-averse lender would demand an extra premium from group A members because var\((d^*/|Y_1^*\)) > var\((d^*/|Y_2^*\)). Therefore, a risk-averse lender will offer more favorable terms to those members of group B with \( Y_1 = Y_1 > Y_2 \). Once again, although this situation is consistent with credit being supplied by unprejudiced, profit-maximizing lenders, many observers might regard such differential treatment as discriminatory since the mean default rates are equal.
If regulation limited the information set to \( Y \), the lender would again estimate a single default loss equation, \( AB \), using pooled data from groups \( A \) and \( B \). A risk-neutral lender basing decisions on default loss predictions from \( AB \) would provide less (more) favorable credit terms to members of group \( A \) (\( B \)) for whom \( Y'_r > Y'_l > Y \). The relative advantage is reserved for \( Y'_r > Y'_l < Y \). A risk-averse lender would react to the fact that \( \text{var}(d[Y]) > \text{var}(d[Y_l]) > \text{var}(d[Y_l]) \). Thus, in addition to the effects discussed above, restricting the optimal information set would induce a risk-averse lender to further reduce credit supply to group \( B \) and simultaneously to increase supply to group \( A \).

Thus far, our analysis has assumed that lenders do not alter the information set in response to regulation. If a regulation bans the use of \( Y \) or \( Y_l \), lenders would recompute the marginal costs and benefits of adding an additional variable, \( Y \), and might decide that such an expansion of the information set is justified. A formal model of such adjustments is beyond the scope of this paper. However, in the next section we present some empirical evidence on the consequences of making such an adjustment.

IV. EMPIRICAL ILLUSTRATION OF SOME EFFECTS OF INFORMATIONAL RESTRICTIONS

The previous section indicates that the effects of restricting the information set available to lenders cannot be determined from theory alone. One must know something about the distribution of all variables in the information set before and after regulation. Moreover, when several variables, measured with error, are used to estimate the default loss equation, it is not generally possible to determine a priori the effects of restricting information.

This section presents an empirical illustration of the potential effects on the mortgage lending decision of regulation limiting use of information regarding property location. Such limits are implicit in the Community Reinvestment Act which is one of several important and recent regulations designed to influence the information set used by lenders.

The first step in this empirical illustration is the estimation of a series of alternative default loss equations using different information sets. The data used for this purpose are taken from the FHA section 203(b) mortgage insurance program and consist of 8,618 mortgages written in 1974 by ten SMSAs. Default losses on these insured mortgages are observed over the 1974-1980 period. These mortgages represent a fairly standard loan product. All have a thirty year term, the loan amounts range from only $15,000 to $30,000, and the other terms are standardized under the section 203(b) insurance program.

Unlike the estimates reported in a number of previous studies reviewed by Barth, Cordes and Weaver (1979, 1980, 1983), the empirical default loss estimates reported here are based upon a regression equation in which the actual dollar default loss is the dependent variable. Thus our results are not subject to criticisms made in Avery (1978) and Eisenberg (1977) of simple binary default models in which the dependent variable is unity in the case of default and zero otherwise. An additional advantage of the section 203(b) mortgage data employed here is the limitation placed on lender screening because the FHA may not use property location in making its insurance decision. As a result, location should not be a criterion for sample selection in the mortgage data used here.

8. The impact of including an alternate variable \( Y \), depends on a complex way on the partial correlation between \( Y \) and both \( Y_r \) and \( Y_l \). These issues are discussed in Medoff (1977).
In our empirical illustration, the information variable \( Y \) is assumed to be a vector of characteristics including: \( MPAY \), monthly payment to income ratio; \( BLDAGE \), structure age; \( DARATIO \), ratio of debt to debt plus assets of borrower; \( LOANVR \), loan to value ratio; \( FHAVLU \), assessed value of structure; \( BRICK \), a dummy variable equal to 1 for masonry structures and zero otherwise; and \( DETACH \), a dummy variable equal to 1 for detached structures and zero otherwise. The \( Y \) variable is property location, central city or suburb. Some elements of \( Y \) are observed with error, but \( Y \), which is discouraged as a lending criterion in the Community Reinvestment Act, is observed without error. Other borrower characteristics which are prohibited under the Equal Credit Opportunity Act are excluded.

Columns A1 and A2 of table 1 are the default loss equations which the lender

<table>
<thead>
<tr>
<th>Variable</th>
<th>City Only A1</th>
<th>Suburban Only A2</th>
<th>Pooled Data B1</th>
<th>Pooled Data C1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-1550</td>
<td>-744</td>
<td>-1297</td>
<td>-1294</td>
</tr>
<tr>
<td></td>
<td>(-3.15)</td>
<td>(-1.49)</td>
<td>(-3.41)</td>
<td>(-3.50)</td>
</tr>
<tr>
<td>( MPAY )</td>
<td>962</td>
<td>951</td>
<td>886</td>
<td>878</td>
</tr>
<tr>
<td></td>
<td>(1.88)</td>
<td>(1.84)</td>
<td>(2.45)</td>
<td>(2.42)</td>
</tr>
<tr>
<td>( BLDAGE )</td>
<td>-3.4</td>
<td>-7.9</td>
<td>-4.7</td>
<td>-5.1</td>
</tr>
<tr>
<td></td>
<td>(-1.71)</td>
<td>(-3.97)</td>
<td>(-2.98)</td>
<td>(-3.23)</td>
</tr>
<tr>
<td>( DARATIO )</td>
<td>254</td>
<td>326</td>
<td>290</td>
<td>293</td>
</tr>
<tr>
<td></td>
<td>(2.87)</td>
<td>(4.27)</td>
<td>(4.74)</td>
<td>(4.73)</td>
</tr>
<tr>
<td>( LOANVR )</td>
<td>17.8</td>
<td>9.1</td>
<td>14.4</td>
<td>14.3</td>
</tr>
<tr>
<td></td>
<td>(3.64)</td>
<td>(1.87)</td>
<td>(4.09)</td>
<td>(4.07)</td>
</tr>
<tr>
<td>( FHAVLU )</td>
<td>-0.009</td>
<td>-0.013</td>
<td>-0.011</td>
<td>-0.010</td>
</tr>
<tr>
<td></td>
<td>(-2.31)</td>
<td>(-2.97)</td>
<td>(-4.06)</td>
<td>(-3.62)</td>
</tr>
<tr>
<td>( BRICK )</td>
<td>135</td>
<td>170</td>
<td>153</td>
<td>156</td>
</tr>
<tr>
<td></td>
<td>(2.16)</td>
<td>(2.58)</td>
<td>(3.34)</td>
<td>(3.43)</td>
</tr>
<tr>
<td>( DETACH )</td>
<td>259</td>
<td>256</td>
<td>229</td>
<td>216</td>
</tr>
<tr>
<td></td>
<td>(1.95)</td>
<td>(3.03)</td>
<td>(3.44)</td>
<td>(3.21)</td>
</tr>
<tr>
<td>( FAIR )</td>
<td>224</td>
<td>(2.86)</td>
<td>44.6</td>
<td>(0.82)</td>
</tr>
<tr>
<td>( GOOD )</td>
<td>5301</td>
<td>3317</td>
<td>8618</td>
<td>8618</td>
</tr>
<tr>
<td>Std. Error of Reg.</td>
<td>1810</td>
<td>1905</td>
<td>1699</td>
<td>1698</td>
</tr>
<tr>
<td>Mean of Dep. Var.</td>
<td>348</td>
<td>226</td>
<td>308</td>
<td>302</td>
</tr>
</tbody>
</table>
would use to classify applicants in the absence of regulation. If information on property location were prohibited, the lender would not be able to group applications by location, and hence would have to base decisions on estimates obtained from a pooled regression. Estimation of such a regression with only $Y$ produces the results displayed in column B1. Lenders might also react by expanding the information set to include variables not initially in the optimal information set due to their high cost and/or low marginal benefit. An example of such a variable is property condition as determined by an inspection. Column C1 of table 1 presents an estimated default loss equation which includes: \( FAIR \), a dummy variable equal to 1 for structures in fair condition and zero otherwise; and \( GOOD \), a dummy variable equal to 1 for structures in good condition and zero otherwise. The reference category for property condition is therefore structures evaluated as excellent. The estimates in table 1 are generally consistent with the literature on default equations in that increasing \( MVX \), \( DARRATIO \), and \( LOANVR \) tends to raise expected default loss, as does inferior structure condition. 9

Table 2 presents summary statistics which describe the impact of limiting the use of information about property location. The first row of table 2 shows that using the pooled default loss equation B1 reduces the estimated default loss of 27 percent of suburban applicants, as compared to 72 percent of center city residents. Moreover, the standard error of the regression for model A1 (A2) is larger (smaller) than that for model B1. Thus both the mean and variance effects of the information restriction tend to benefit center city borrowers at the expense of suburban borrowers.

### TABLE 2

<table>
<thead>
<tr>
<th>Suburb</th>
<th>General City</th>
</tr>
</thead>
<tbody>
<tr>
<td>E (Loss) Model B — E (Loss) Model A</td>
<td>+ 834</td>
</tr>
<tr>
<td>Downpayment Model B — Downpayment Model A</td>
<td>+ 1728</td>
</tr>
<tr>
<td>Model A Default &gt; Model C Default</td>
<td>25.4%</td>
</tr>
<tr>
<td>E (Loss) Model C — E (Loss) Model A</td>
<td>+ 48</td>
</tr>
<tr>
<td>Downpayment Model C — Downpayment Model A</td>
<td>+ 1523</td>
</tr>
</tbody>
</table>

The second and third rows of table 2 present estimates of the econometric effects of such regulation-induced reclassification of applicants. The second row indicates that permitting the lender to use only the pooled model B lowers the lender's estimate of the default loss by $14 (7 percent of the $348 mean) in the case of urban applicants with $Y$, variables set equal to their sample means, and raises the estimated loss by $54

9. Given that default loss is truncated at zero, the estimated coefficients in table 1 may differ from the true structural parameters by a proportional truncation bias as discussed in Poster and Melin (1978). Alternative Tobit estimates of the default loss equations were obtained but produced similar empirical results. In any event, the OLS results in table 1 most likely reflect the estimation procedure which lenders might actually use.
(24 percent of the $285 mean) for suburban applicants with comparable characteristics. Since FHA mortgage interest rates do not vary among applicants, we are not able to estimate the impact of suppressing information about location on the interest rate charged on urban as opposed to suburban borrowers. However, the loan-to-value ratio required of different applicants can vary. Hence, the fixed row of table 2 shows the impact of restricting information on the loan-to-value ratio. In the pooled model, an applicant with Y characteristics equal to the mean of the FHA data would have an expected default loss of $308 if LOANVR equalled its pooled sample mean of 92.9 percent of the property values. If the lender had been able to classify applicants by property location, the same expected default loss would be obtained on an urban mortgage if the loan were 92.1 percent of the property value, and on a suburban mortgage if the loan were 98.8 percent of the property value. Since the mean property value in the sample is $29,585, these changes in LOANVR translate into a $234 reduction in the downpayment required of urban borrowers and a $1728 increase in the downpayment required of suburban borrowers.

Results in the last three rows of table 2 show that the addition of extra variables to the pooled regression in response to an informational restriction does not change the relative impact on suburban and center city mortgage applicants. However, in calculating the fixed cost effects on city and suburban borrowers, one would also have to consider the increase in information costs due to the collection of the extra information on property condition to estimate model C. Indeed, in such a case restricting the information set might have little favorable impact on the final credit terms offered to center city borrowers.

The results presented in tables 1 and 2 would change with variation in the information set or the mortgage portfolio forming the basis for the default loss equation. Also, the 1974-1980 period sample spanned by the FHA data was at a time of rising real housing prices and low rates of default loss. Default losses would surely be larger in a weak housing market. Moreover, the results do not consider default losses after the sixth year and do not discount the observed losses to present value at the time of loan endorsement. Thus, the dollar impacts estimated above may underestimate expectations of possible future default losses.

V. SUMMARY AND CONCLUSIONS

Our theoretical analysis demonstrates that informational restrictions on credit scoring schemes used by lenders may produce undesirable credit supply responses. The impact of such informational restrictions depends on the information set currently being used by a lender. Lenders may also respond to regulation by expanding the information set, raising costs of screening applicants. Moreover, it is difficult to distinguish and define discriminatory lending. Lender behavior that appears to be discriminatory may actually be consistent with non-discriminating profit maximization when information on borrower characteristics is costly and imperfect.

Empirical analysis of an actual regulation which limits the use of property location in mortgage lending decisions illustrates how regulating credit scoring schemes can favor a particular group—city residents—at the expense of another group—suburban homeowners. However, the quantitative impact of such favorable treatment of center city applicants appears to be relatively modest, particularly if the restriction on information induces lenders to expand the information set, raising the cost of information to all borrowers. This finding may reflect the relatively strong
housing market which prevailed during the period covered by the FHA data. In any event, our empirical analysis illustrates the potential pitfalls to be avoided when designing new regulations.

REFERENCES


