A Shift in the Mortgage Landscape: The 1990s Move to Automated Credit Evaluations

John W. Straka*

Abstract

This article reviews the shift in the 1990s from traditional manual mortgage underwriting to mortgage scoring models and statistical automated underwriting (AU). AU has become the predominant mortgage underwriting method, helping to catalyze a broad wave of technological advancement. Still, further significant changes are likely.

Examining statistical AU’s history, growth, and underlying factors and concerns, this article especially considers the question of how far AU should be relied upon to evaluate more marginal and higher-risk mortgage applicants (traditionally harder cases). Supporting a cautious, continued expansion of statistical AU into higher-risk loans, the article reveals some provocative evidence on statistical scorecards versus traditional underwriters and explores fair lending strengths of AU. Future overall prospects for likely AU and related research areas and developments are summarized.

Keywords: Automated underwriting; Credit scores; Mortgage scoring

Introduction

The evolution of housing finance changed in the 1990s (especially after 1995) with the growth of automated statistical credit and mortgage scoring as a method for underwriting and approving loans. Automated underwriting (AU), previously used in credit card and auto lending but largely absent from the mortgage business—and typically launched or expanded with customized application scoring models—has now grown into the predominant (around 60 to 70 percent and growing) mortgage underwriting method. Risk- or value-based pricing seems poised to move toward finer levels of automated risk distinctions at the loan level. Scoring or other statistical models and rapid automated decision tools have similarly grown in mortgage servicing, collateral assessment, portfolio analysis, and other areas. Coupled with technology, the mortgage industry’s reliance on statistical decision tools has come a long way since empirical mortgage scoring models began to come into use roughly one decade ago. Many observers believe there is still more to come.1

This article reviews the role of borrower credit history and traditional mortgage credit risk underwriting—assessing overall expected mortgage default risk—and examines how the insti-

* John W. Straka is Director of Consumer Modeling at Freddie Mac.

The author thanks Peter Zorn, Kevin Winkler, Jeff Markowitz, Peter Mahoney, Tom Holloway, Jim Follain, Amy Bogdon, and two anonymous referees for comments and suggestions. This is a revised version of a paper presented at the American Real Estate and Urban Economics Association Mid-Year Meeting, May 30 to 31, 2000, in Washington, DC. The views expressed in this article are those of the author, and do not necessarily represent the views of Freddie Mac or its Board of Directors.

1 For just one example (a summary of a recent mortgage technology conference), see Brockman (2000).
tutional nature of this risk assessment changed in the 1990s.\(^2\) The automation of credit risk evaluation helped to catalyze a modeling and technological revolution that has been reshaping the landscape of risk evaluation, origination, and loan processing in the mortgage business (Strickberger 1999b). An underwriting and origination system of manual, decentralized, labor- and paper-intensive loan processing and risk assessment over a matter of weeks (or even months) has shifted increasingly to more centralized and streamlined model-based risk evaluations that are heavily automated, with conditional AU decisions in minutes (and same-day closings in some cases). This type of automation substitution and growth has become a pattern, with credit data processing and other evaluations now shifting increasingly to the Internet (Cornwell 1999), especially impacting business-to-business mortgage industry commerce.\(^3\)

These developments, of course, have not happened overnight—nor without new needs, costs, issues, pitfalls, and detractors—and they continue. This article has six sections. The first section briefly reviews the state of mortgage underwriting in the 1980s and early 1990s. The second section describes some of the first predecessors, or first wave, in the early 1990s of the new AU age. The third section briefly reviews scoring model types and methods in the context of their growing industry developments in the early and mid-1990s. The fourth section reviews the relatively rapid industry adoption of statistical mortgage scoring models and AU that began in 1994 to 1995. The fifth (and longest) section discusses the economic incentives driving and expanding these changes and explores various concerns and issues. It shows evidence suggesting that, while not a panacea, statistical AU tends to be not just faster but also more efficient, consistent, objective, and accurate than typical mortgage underwriting of the past. Customized statistical AU may better assess credit risk even (and perhaps especially) on gray-area marginal and higher-risk loans. Further, it may be able to qualify a material share of traditional subprime borrowers for prime-loan (or near-prime) pricing and may also raise, on balance, lower-income and minority loan volumes in comparison with historical underwriting, while also bringing other fair lending advantages.

The potential to eradicate discrimination from mortgage underwriting, to deny no applicants, and to bring competitive pricing that is fair, objective, statistically based, increasingly lower-cost, and automated, for all risk groups is one of the greatest opportunities presented by statistical AU and its developing antecedents. Marginal and higher-risk loans are traditionally more costly to underwrite (and service) and harder to automate, but some institutions have sought to better model these risks and automate their assessment, which may avoid a “digital divide.” The coming decade will better tell if this is an aim beyond feasible proportions or an achievable vision.\(^4\)

Reliance on statistical models, AU, risk-based pricing, and so forth, with their many demands and challenges, seems likely to expand further in the decade ahead, driven by competitive

---

\(^2\) Other risks, such as prepayment, loss severity, and capital risk, will be implied or referenced (e.g., in connection with risk-based pricing) but are not a main focus of this review.

\(^3\) See the article by Michael LaCour-Little in this issue for a broad review of the evolution of technological change across mortgage finance. Point-of-sale Internet AU is also increasingly common.

\(^4\) Some loans cannot be evaluated by AU, of course, but those with missing or inadequate data, for example, are a comparatively small percentage. Freddie Mac and Fannie Mae, along with other institutions, have in recent years been on a mission to increasingly compete across markets and further develop and extend their respective AU and other risk-evaluation systems. The Freddie Mac and Fannie Mae AU systems also have enhanced liquidity (by promoting greater credit risk standardization) and accelerated broad industry diffusion of AU technology.
industry incentives to further reduce origination costs and expand all markets. Industry research, education, and resource needs to study the development, uses, data inputs, and benefits/costs of statistical AU are likely to grow along with loan-level costing/pricing models and other statistical models in related applications. The final section of this article explores some likely future developments and research areas.

**Mortgage Underwriting in the 1980s and Early 1990s**

Historically, for most mortgages, manual underwriters have played a central role by reviewing the credit report and the rest of a painstakingly assembled loan file to render a judgment about the risk of each loan. From 1982 to the mid-1990s, the secondary market relied almost exclusively on delegated underwriting, in which thousands of lenders with tens of thousands of mortgage underwriters followed their own written standards and the voluminous guidelines of Fannie Mae or Freddie Mac (or those of similarly functioning counterparts for non-conforming loans).\(^5\)

Lenders and investors require such underwriting in good measure because, just as in other forms of consumer finance, the mortgage borrower’s credit record is a very important risk predictor. Borrowers with poor credit records go into mortgage default at much higher rates than do borrowers with unblemished or good credit. Mortgage underwriting, however, is relatively more complex than credit card or typical auto-loan underwriting (e.g., because of the importance of equity), with larger loans at risk of default or fraud, and thus it has historically been viewed as less readily automated. Mortgages also have a longer time profile of performance (scoring models for credit cards can be built and updated with just two years, or less, of performance data).

Empirical mortgage default research through the late 1980s and early 1990s had discovered that the major empirical driver of default (at least real estate–owned) risk is negative-equity risk, reflected in original and mark-to-market loan-to-value ratio (LTV) (e.g., Foster and Van Order 1984). Virtually none of this default modeling (at least none in the public domain) had been able to include borrower credit records. Although mortgage underwriters, managers, and investors recognized the importance of credit history (approximately half of mortgage denials have been for reasons of credit), and generic credit analysts knew of its general importance, its relative empirical importance for mortgages was not well understood. Belief in the possibility of “equity underwriting” (using only, or mostly, LTV) helped spawn some ill-fated industry experiments, beginning in the latter 1980s, which became known as the “low doc era.” Citicorp, for example, suffered large mortgage losses at the time (and other setbacks), nearly becoming insolvent.

By the early 1990s, decentralized manual mortgage underwriting, with a trained and guided but ultimately subjective evaluation of each loan, remained the dominant and nearly universal procedure for assessing individual loan credit risk and approving or denying most mortgage applications. Mortgage model builders, on the default risk side, were mostly absorbed

---

\(^5\) Prior to 1982, with much smaller loan volumes than now, Fannie Mae and Freddie Mac each employed a staff of underwriters to underwrite (i.e., re-underwrite) every single loan file for each loan purchased. Each then switched to postfunding quality-control review of a sample of loans, with lenders contractually assuming the chief responsibility for individual-loan credit risk evaluations.
with pool-level models, costing simulations, and the study of regional house price cycles and negative-equity risks. This bifurcated system of credit risk evaluation would soon change and quite rapidly begin to evolve in a new direction.

The First Signs of Change

The traditional “three Cs” of mortgage underwriting are collateral, capacity, and credit. Collateral is assessed through down payment and the appraisal (or LTV). Capacity is measured through variables such as debt-to-income ratio, employment status, and borrower reserves. Credit is assessed through the details of the borrower’s credit report. The use of these variables in traditional manual guideline underwriting was based on business experience—poor performance of loans with high LTVs, for example (especially in soft markets). By the early 1990s, because sufficient credit data was almost never available (even within large mortgage institutions) and because of its importance, credit was the key missing piece of the three Cs puzzle for mortgage default modelers.

Interest in bringing credit scoring methods into mortgage underwriting and finance is quite old within the mortgage business. Anecdotal reports of experiments with “mortgage scorecards” extend back several decades. The first institutions to develop modern statistically based mortgage scoring models included PMI Mortgage Insurance Company (the pmiAURA model, beginning in the late 1980s) and Citicorp (in 1992). These pioneering early models, while not widely discussed, gained the first footholds in the mortgage business for mortgage scoring. More comprehensive industry acceptance of statistical loan scoring technology would soon begin to unfold.

One broader development began in 1992 when Freddie Mac completed its first empirical study of generic credit scores (using FICO [Fair, Isaac and Company] bureau scores) and their feasibility for predicting mortgage default (Holloway, MacDonald, and Straka [1993] summarized results from this internal study). A main conclusion was that scores worked as a statistically significant and strong predictor in a mortgage default equation (based on Freddie Mac loan data), suggesting that it was feasible to use them as a component of a mortgage scoring model. As unremarkable as this now seems, all modelers did not easily accept it.

Despite widespread mortgage underwriting practice and the common use of credit scoring in other forms of consumer finance, a pure option-theoretic view of mortgage default could hold that variables other than negative-equity risk measures (LTV measures primarily) should not matter. The credit score finding was more consistent with earlier lines of default modeling and with a broader consumer-choice view of mortgage default (Quercia and Stegman 1992; Vandell and Thibodeau 1985; Zorn and Lea 1989). Theoretical objections from the options perspective were less clear because mark-to-market LTV cannot be measured at the individual-home level, and other evidence pointed away from the pure option-theory view (Lekkas, Quigley, and Van Order 1993; Quigley and Van Order 1995).

---

6 This is not meant to be an exhaustive list. Other relatively early model developers reportedly included Prudential, Bank of America, American Savings, and at least one or two finance companies who had models of some kind. There is an element of business secrecy here, since some institutions sought to gain loan trading and competitive origination advantages. AU for mortgages was already well established by 1995 in Australia (Neagle 1995).
With more detailed loan-level data frequently lacking, it was especially easy for modelers focused on the strong default to LTV (original and current) relationship to neglect the significance of other variables, including credit. Despite the strong LTV relationship, most borrowers with negative equity do not default, and credit history at origination is a good predictor of relative default likelihood in both weak and strong housing markets, a finding that has now been confirmed through multiple stages of modeling and research at Freddie Mac and elsewhere. This is illustrated in figure 1, which summarizes a 1995 assessment of 1990 to 1991 Freddie Mac loans using FICO credit scores and data from a weak regional housing market and a strong regional market. Findings like this have demonstrated that option-theoretic and consumer-choice views are complementary rather than competing explanations of mortgage default.7

Along with its credit score findings during this era, Freddie Mac also confirmed the significance and robustness of several noncredit underwriting variables—including, of course, LTV—in loan-level default models estimated on large databases from Freddie Mac, mortgage insurers, lenders, and other institutions. Based on this research, development of Freddie Mac’s mortgage scoring models and AU system, subsequently named Loan Prospector (LP), began in earnest.8 Throughout this article, cited Freddie Mac experiences should be regarded as merely examples—or windows on the broad industry movement that adopted AU and mortgage scoring models during the past decade.9

Scoring Model Types and Methods

The finding that credit bureau scores are strong predictors of mortgage default was somewhat surprising because these scores are generic—designed to be predictive of consumer default risk, using large nationwide samples of credit data alone (e.g., the now well-known FICO score is most predictive of 90 or more days delinquency on an account).10 Custom models, such as a mortgage scoring model, are designed to be predictive of default risk on a specific type

---

7 Borrowers with negative equity and poor credit records may be, by far, the most likely to default, in good measure because they have neither equity nor good credit to preserve. Most borrowers seek to maintain good credit, and generally succeed, while many with poor credit are unable to recover or to build good credit very rapidly. For these reasons at least, credit history at origination (similar to original LTV versus mark-to-market) tends to be a good predictor of subsequent credit, the default incentive, and actual default behavior. More extensive theoretical and empirical discussion would require another article. See, for example, Elmer and Seelig (1999) and Yang, Buist, and Megbolugbe (1998).

8 Led by Henry Cassidy, the original “scorecard team” included Tom Holloway, Gregor MacDonald, Andrew Jaske, Michael Bradley, this author, Doug Gordon, Larry Cordell, Doug McManus, and Betsy Bird, supported by strong analysts (Molly Boesel, Clayton Botkin, Brian Broomfield, Joe Emmim, John Mulligan, Connie Soper, and others) and consultations from leading economists, underwriters, and legal experts (in particular Peter Mahoney) across Freddie Mac, as well as scoring-industry consultants. The development and implementation of LP overall, plus many later enhancements, would come to be a feat achieved by a dedicated sizable team at Freddie Mac, too numerous to name here.

9 Freddie Mac’s experiences are most directly familiar to the author. Specifics of most AU systems in the industry have tended to be quite proprietary, and Freddie Mac’s adoption of AU and mortgage scoring, in the historical record, did have significant industry influence. None of this is meant to imply that Freddie Mac was the first developer of mortgage scoring models (as previously noted) or that Freddie Mac was the center of this modeling and AU universe (wherever that might be).

10 In both the modeled default events and the predictor variables of a generic score, mortgages are treated the same as all other credit trade lines.
Figure 1. Default Rates by Credit Score (FICO) in Weak and Strong Markets, 1990 to 1991 Originations Sample, Performance through May 1995

Note: Using a sample of Freddie Mac loans, this figure illustrates the relationship between credit scores, default, and markets with weak and strong house price change. Sample sizes for >=760, 700–759, 620–699, and <620 score, respectively, were Weak Market >=90: 1,202, 2,636, 2,058, 447; Weak Market Total: 25,204, 36,098, 27,070, 5,728; Strong Market >=90: 640, 1,266, 758, 149; Strong Market Total: 7,528, 9,336, 6,018, 1,423.
of consumer credit and/or a specific institution’s customer base and can include all data from the loan application: credit, LTV, product type, property type, debt ratios, and other economic information.

On the credit side, a custom model may use the FICO score (or some other score) alone, a credit score in conjunction with several detailed credit characteristics (such as recent mortgage lates), or only detailed credit characteristics. A partially customized model may be limited in various respects; for example, mortgage scores designed to predict mortgage default but limited to credit data only were developed and marketed in the mid-1990s (e.g., Equifax’s Mortgage Score). A fully customized model, in contrast, will consider a complete range of detailed credit characteristics, credit scores, and all relevant noncredit data, resulting in a fully optimized model for the particular consumer application, which may include various segmentations or interactions. No score by itself is sufficient for any AU system or other complete business application. For further discussion, see Holloway and Jaske (1996) and Makuch (1998a, 1998b).

The quality of a mortgage scoring or similar statistical model, of course, is in good measure determined by the quality of the data used for development and testing. Developing good modeling data is not nearly so simple as obtaining a “data dump” from a financial institution’s records. Besides sampling, availability, and storage-form issues, various data fields raise issues that require decisions. Most institutions and modelers have experienced a learning curve regarding their loan-level data; data coded from loan files and from new or external sources can be particularly challenging to develop. In a very important development, individual loan performance data from mortgage servicing became enhanced, more widely available, and more standard in the 1980s and early 1990s, a critically necessary ingredient for empirical loan-level modeling.

Virtually no institution was storing credit records on mortgage loans in an easily accessible medium in the early 1990s. Major investors and conduits did not require this, and most institutions also had little or no experience with the steps required to obtain historical credit data electronically from the major credit repositories’ archive records. By the mid-1990s, especially among larger institutions, this was rapidly changing.11 The complexities of credit data lie beyond the scope of this article, but acquisition and development of good credit data for modeling and implementation requires substantial expertise. Overall, the development, testing, and interpretation of good modeling data has generally been by far the most time- and resource-consuming part of a typical scoring model development.

Mortgage scoring models can be derived using any one or several of various modeling techniques.12 All of these modeling techniques as a whole are statistically quite general and have been applied to a broad array of studies and uses. Mortgage scoring models are distinguished by their application of domain expertise—that is, the business, data, and institutional exper-

---

11 Improvements both in electronic storage of current records and the coverage (borrower match rates) of repository historical records joined with growing awareness of the importance of credit data.

12 Makuch (1998b, 69–74) describes well and classifies these modeling technologies into methods: regression-based, neural networks, tree-based, and segmentation. A new wrinkle in mortgage modeling compared with other applications of statistical scoring systems is that hazard regression technologies are increasingly used, reflecting the importance of prepay terminations, longer time profiles of cash flows, and so forth.
tise of the model builders—and their use of scorecard technology (for a historical description of scoring methods and best practices, see Lewis 1992).

A statistical mortgage underwriting scorecard is nothing more than a particular mathematical transformation of a mortgage default model's coefficients. A scorecard is not strictly necessary to deploy this type of model for its intended application, but scorecards are very useful communication, data processing, and information-summary tools, particularly in model deployment and monitoring and risk management.

Industry Adoption of Mortgage Scoring and AU

AU was still quite novel and obscure in the mortgage industry in the spring of 1994 when Freddie Mac deployed, with a group of lender partners, the first pilot version of its AU service, LP. Meanwhile, system developers at Fannie Mae and elsewhere were building or refining their own AU systems, but some had initially chosen a different path from statistical mortgage scoring: mentored artificial intelligence (AI).

In a mentored AI AU system, AI algorithms are used to try to train a system to reproduce the credit decisions of a human underwriter (or group). Manual guideline or rules-based underwriting algorithms may also be used. Makuch (1998a, 4) refers to systems of this kind that “for the most part came and went” during the early to mid-1990s. They brought speed and consistency to the underwriting process but were not designed to optimally predict defaults. These systems included Fannie Mae’s first versions of its Desktop Underwriter (DU) AU system, and General Electric (GE) Capital’s first AU system (GENIUS). By the end of 1995, mentored AI systems had largely lost out to or begun to progress to statistical mortgage scoring—which brought the key advantage of modeling the actual likelihood of mortgage default. 13

Another prominent question was whether AU systems would be capable of handling special affordable loans that were more often found to carry higher risk and, more generally, all higher risk marginal or harder-case gray-area loans. In September 1994, Freddie Mac published (after pilot testing) its Gold Measure Worksheet (GMW). This was a statistically based manual mortgage scoring worksheet for noncredit and credit risk factors (with a FICO bureau score, MDS [CNN-MDS, Inc.] bankruptcy score, or detailed credit alternative), designed to assist underwriters in evaluating risk layering, especially on affordable loans that were stretching the limits of traditional underwriting. 14

13 GE switched toward mortgage credit scoring, and Fannie Mae soon switched roads also to pursue the same mortgage scoring model path as Freddie Mac’s LP, the pmiAURA model, and others (Inside Mortgage Finance 1995a, 1995c, 1995g, and 1996a). DU implemented its first statistical scoring model in the spring of 1997. Legally, business defensibility also favors statistical scoring (discussed more below). Mentored AI systems still in use, or more recently developed (e.g., rules-based for subprime lending), may have relatively short half-lives remaining.

14 An automated personal computer–based version was also released in the spring of 1995 (Inside Mortgage Finance 1995d). The GMW (for a copy, see Avery et al. [1996, 643]) also demonstrated the process of scoring (in effect providing a unique, early peek-under-the-hood into broader developing AU systems such as LP). One year later, Freddie Mac began offering to process its Affordable Gold loans through LP, and Fannie Mae did likewise with its Community Homebuyer affordable products through DU.
Nineteen ninety-five saw many developments; among them was licensed use of the earlier developed pmiAURA AU system. In February, Freddie Mac rolled out the first full commercial version of LP, and two months later Fannie Mae followed with its first release of DU (which included a jumbo-loan capability from Residential Funding Corporation). It quickly became clear that, through their broad availability, the Freddie Mac and Fannie Mae AU systems would tend to “level the playing field” between larger, more technologically sophisticated lenders and midsize to smaller lenders. The U.S. Department of Housing and Urban Development (HUD) began reviewing and approving lenders’ AU systems for use on Federal Housing Administration (FHA) loans, beginning with a system developed by Countrywide (CLUES) and then pmiAURA. Private mortgage insurers backed the Freddie Mac and Fannie Mae systems, since many had been developing or planning their own AU, such as GE Capital Mortgage Insurance’s Omniscore and Mortgage Guaranty Insurance Corporation’s Mortgage Score. All of these systems helped to spur technological and policy efforts to reengineer and streamline the origination process, and discussions of uniform industry technology standards began to emerge. AU’s widespread adoption was underway.

Of course, traditional manual underwriting of loans was still widespread—but this also moved toward scoring. In July 1995, Freddie Mac published an industry letter that endorsed the use of FICO or MDS scores in helping to assess credit quality. The letter also notified lenders that Freddie Mac had begun using credit scores in its underwriting quality control. Fannie Mae followed three months later with a similar industry letter (Kulkosky 1995). GE and others also endorsed credit scores. In subsequent years, FICO/LTV matrices, for example (typically with empirical backing of some kind), have become increasingly common underwriting tools.

On Wall Street, the ratings agencies moved to endorse mortgage scoring. Standard & Poor’s allied with Freddie Mac in 1995 to 1996 to release a risk grading system, attached to LP, for jumbo and subprime mortgages. The ratings agencies acquired more data and validated the performance of pmiAURA, LP, DU, and other AU systems (e.g., Standard & Poor’s 1997) as they further developed and refined their own credit risk evaluation systems. PmiAURA became the first mortgage score accepted by all four ratings agencies as an effective tool for establishing levels of credit support needed on securities backed by nonconforming, conventional loans (PMI Mortgage Insurance Company 1996).

Government ratification of the use of credit scores (important for many banks) came when the Federal Reserve Board published its own study of the statistical validity of credit scores in

---

15 Six of the top 12 mortgage lenders used pmiAURA as of December 1996. PmiAURA was (likely) the first AU model to predict default likelihood over the full life of a loan. For the industry as a whole, this began the process of moving AU beyond the shorter-term horizon of traditional mortgage underwriting (generally about two to five years), to align it better with the full-term horizon view of mortgage investors.

16 See, for example, Inside Mortgage Finance (1995b, 1995e, 1995f) and Secondary Marketing Executive (1996) for summaries of these events and systems.

17 Both the Freddie Mac and Fannie Mae industry letters also stressed that while they are highly predictive of default risk, credit scores alone are definitely not sufficient for complete underwriting of any kind. For a good discussion of credit scoring and some of its first broader cultural impacts on the mortgage credit business, see Sonntag (1995).

18 Subprime (“B” and “C” rated) mortgages have been an area of relatively slower diffusion of scoring models, although scores of different kinds are increasingly used (Cornwell 1997; National Mortgage News 1999).
predicting mortgage default (Avery et al. 1996). Custom FHA and U.S. Department of Veterans Affairs (VA) models were built by Freddie Mac, in work for those agencies, and incorporated into LP, gaining final approval in 1997 and 1998. DU and other AU systems were later approved for customized FHA and VA use also, and the FHA has now completed its own mortgage scoring model, expected to be implemented soon.19

Today most mortgage applications (likely almost all for conventional loans) are scored in some sense (at least a FICO score is used, for example), although uses vary. Approximately 60 to 70 percent of all mortgages are now underwritten by an AU system, and this share continues to rise.20 Besides mortgage scoring of applications in AU, other industry statistical and scoring tools have subsequently been developed. Behavior scoring models for collections, servicing, and loss mitigation have gained an increasingly visible role, with most mortgage servicing operations now using these sorts of tools (Cohen 1997; Cordell and Trianna 1999; Inside Mortgage Finance 1996d; Jaske 1997).21

**Underlying Factors and Concerns in the Growth of Mortgage Scoring and AU**

Mortgage AU has grown in response to economic incentives. First, competition and decreased profit margins have brought demands to cut loan origination costs. Assessing implications for the United States, Neagle (1995, 16–17) described established AU for mortgages in Australia: “Origination costs are much lower...there are far fewer underwriters and closers...incremental costs are practically nil.” AU can substantially reduce time spent on most applications, with less fallout, lower hedging costs, and fewer repurchases (Grayson 1996). Other efficiencies, such as eliminating extraneous documentation, have followed. As with many new technologies, the full cost savings were not realized immediately. As of September 1996, Freddie Mac (1996, i), for example, estimated that LP was reducing costs by $300 to $650 per loan, with savings expected to grow (based on lenders’ reports). Changes since then include reductions in the pricing of AU decisions, expansion in the percentage of loans classified as “accepts” (e.g., LP eliminated the class “refers” in 1998), open-system architectures, and point-of-sale AU over the Internet. Lenders have varied in their speed and means to take advantage, but a recent KPMG Consulting survey found in 1999: (1) 27 percent lower underwriting costs and 15 percent lower back-office costs per loan with higher AU usage (comparing usage of over versus under 60 percent), (2) 45 percent higher operational profitability with shorter turn-around times from application to closing (comparing turnaround times of less than versus more than 45 days), and (3) retail channel gains of a one-third lower fallout rate,

---

19 See Inside Mortgage Finance (2000b). Regulators have cautiously endorsed mortgage scoring, recognizing its many benefits for investors, lenders, and consumers, but also pointing out that scoring models and AU systems must be properly developed, implemented, and maintained to comply with fair lending and other regulations.

20 In a recent survey, KPMG Consulting (2000, IV-4) found 60 percent of total loans underwritten by AU in 1999. More recent observations suggest that this figure is now higher.

21 In the behavioral models, credit data can usually be scored in the model in some fashion, but once a borrower has entered a delinquent state (and particularly serious delinquency), mark-to-market LTV typically has the large role in determining the likelihood of eventual loss.
34 percent lower underwriting costs per loan, and 37 percent higher back-office productivity with higher AU usage (comparing usage of over versus under 60 percent).\footnote{See KPMG Consulting (2000, III-19, IV-9–13). Substitution of AU for the more costly manual process is still underway. The timing and extent of cost savings to consumers depends on competition and market structure, scale economies, technical progress, investment costs, and so forth, and requires more research. Besides the reduced time and hassles for approval, consumers have seen appraisal and other costs come down in a likely growing trend.}

Second, for portfolio lenders and others using statistical scoring and AU, underwriting and quality reviews are now considerably more accurate and efficient. Likewise, for Freddie Mac and Fannie Mae and other purchasers or guarantors, automated scores revealed (in greater clarity and depth) the extent of principal-agent-based adverse selection that had been included in traditional business volumes. Industry default losses have been reduced or better covered by improved screening or pricing of the highest-risk “tail” of loans that had previously been included under manual underwriting and quality control. This high-risk tail was a small percentage of the loans being approved and sold but it accounted for a much higher percentage of actual and expected default losses. Delegated underwriting and other buyer-seller (e.g., lender-broker) contracts create many principal-agent-partnering relationships in the mortgage industry. AU has given the principals (purchasers) a clearer, faster, and more efficient means of gauging the agents’ (sellers’) loan qualities. Likewise, the fear of adverse selection from less accurate credit risk evaluation has spurred the growth and diffusion of statistical AU.\footnote{High-risk loans in the market now have far fewer places to go unnoticed for long. Some have suggested that AU provides sellers with a new means to more effectively engage in adverse selection (Passmore and Sparks 1997) and should result in lower interest rates plus reduced profits for securitizers. But a quite costly amount of adverse selection was already taking place up through the mid-1990s. The premise of enhanced adverse selection from AU neglects contractual obligations of sellers and, beyond this, pertains only to the degree that the sellers-agents can discover and effectively use robust statistical relationships or underwriting rules not detectable by the principals-buyers.}

Third, the boom-bust refinance cycles of the 1990s (1992 to 1994 and 1997 to 1999) have clarified for many lenders the cyclical personnel costs of transitory buildups in labor-intensive processes (severance and other fixed costs, lower morale, etc., from layoffs). While other factors affect this also, the industry’s 1999 to 2000 employment bust has been milder than the 1994 to 1995 bust, with total industry employment down 13.2 percent versus 26.9 percent.\footnote{Based on Bureau of Labor Statistics employment (Standard Industry Classification Code 616) through October 2000, preliminary industry employment rose in September for the first time in 17 months but fell in October figures.} AU’s ability to process high cyclical volumes with fewer added staff likely deserves some credit here, and future cycles also seem likely to be diminished, holding all else constant.\footnote{Anecdotal reports suggest that underwriter employment numbers have been hit relatively hard in the 1999 to 2000 contraction, apparently resulting from a combination of flexible contract underwriting and the still-growing shift to automation.}

Fourth, both competitive and regulatory pressures have pushed the industry toward expanding markets, particularly in more affordable (typically higher LTV) loans. Manual guideline underwriting worked around univariate “knockout rules” (such as no loans above a 36 percent debt ratio). Exceptions and risk layering (e.g., poor credit plus debt higher than 36 percent) were allowed (case-by-case or in policy) with little or no risk quantification and feedback. In contrast, scoring tools and AU have allowed the tradeoffs between risk factors to be more precisely quantified, giving the industry greater confidence in “pushing the envelope” of acceptable expected default rates—e.g., enabling Freddie Mac’s entry into 97 percent LTV loans (Inside Mortgage Finance 1996c, 3).
Underwriter Inconsistencies, Gray-Area Loans, and Scores versus Underwriters

Many institutions undertake extensive efforts, trying to make underwriting guidelines and their application fair and consistent. However, imprecision remains, as noted by Freddie Mac (1996, 19):

These efforts, while essential, cannot change the fact that traditional underwriting remains dependent on subjective human judgment...inconsistencies stem from the complex and multifaceted task of tallying up...an underwriting decision ....The cumulative effect of...[lending] mistakes weighs heaviest on households whose applications fall in the gray area between acceptance and denial.

This has heightened fair lending concerns with both potential disparate treatment and disparate impact, highlighted by Home Mortgage Disclosure Act (HMDA) data results on denial disparities in the early 1990s. AU systems have developed and grown partly in response, since they largely dispel disparate-treatment concerns. Sound AU ignores an applicant's minority status and any information not demonstrably relevant to assessing a borrower's statistical likelihood of default. Such consistency and objectivity are long-held underwriting objectives.26

But how far should AU be relied on to evaluate more marginal and higher-risk applicants? One view is expressed by Kunkel (1995, 47–50):

In these cases, human judgment and discretion are needed....Increasingly careful about decisions involving low- and moderate-income lending, most lenders now have a second review committee to examine all denials...A completely automated system...would not be able to function in these situations.

Similarly, many subprime lenders and underwriters have claimed for some time now that their higher-risk manual underwriting cannot be automated through scoring. But the following quote from Lewis (1992, 92) offers a contrasting view that exists among various statistical modelers:

To override the scoring system [at the margin] is the least justifiable policy...It is precisely in the neighborhood of the cut-off that a scoring system is most valuable....Close to the cut-off...scoring systems prove their worth.

Statistical scoring models are similarly making increasing inroads into subprime lending. The Office of the Comptroller of the Currency (OCC) (1997) warned that bank examiners would focus particularly on discretionary manual overrides of scoring systems, scrutinizing such underwriting decisions for disparate treatment.

AU brings a sizable aggregate cost reduction to many lower-risk loans that are easier to rate, but the costs of manually underwriting gray-area and higher-risk applicants are especially large (including reviews for disparate treatment). One important question is whether manual underwriting or statistical AU can better predict default at these higher risks.

26 See Fannie Mae (2000) for a related discussion of historical underwriting practices, AU’s role, differences in minority-applicant risk distributions, and other key facts.
Makuch (1998a, 5) illustrates how an early version of GE’s Omniscore outperformed manual guideline underwriting overall. Three groups, streamline, traditional, and caution, respectively had foreclosure rates of 15, 79, and 471 percent of the total book rate when using the scoring model and 70, 142, and 210 percent when using traditional manual underwriting (with the decline rates in the two methods set equal). Scoring can effectively identify and decline (or price up) more of the appropriate (truly higher-risk) applications and likewise identify and keep (or price) down more of the truly lower-risk applications, reducing both default losses and the social inefficiency of lending errors.

Clear comparisons between scoring models (mortgage or other) and manual underwriters tend to be infrequent, in part because automation gives scoring a major advantage. Once scoring begins within a financial institution, and its usefulness becomes clear, it is also difficult to “put the genie back into the bottle” to conduct natural experiments.27

Before scoring caught on in the mortgage industry, Freddie Mac experienced a natural experiment that created a head-to-head comparison between scores and underwriters—of particular interest because it involved more marginal and higher-risk applicants. The usable statistical sample was more than 700 Affordable Gold loans, from a pool of about 1,000, originated in winter to spring of 1993 to 1994 by a major mortgage lender.28 Performance on these loans provided a “horse race” between laborious manual review underwriters’ ratings and a far faster scoring-based classification.

As shown in figure 2, the race was not very close.29 After three years (a period over which underwriting credit-quality differences are typically strongest), the investment quality loans determined by scores performed much better than the non-investment quality loans determined by scores—scores thus performing as expected. But the investment quality loans determined by the underwriter ratings performed only marginally better, the same, or worse than the non-investment quality loans determined by these ratings. Review underwriter ratings that took six months to complete performed not much better (if better) than flipping coins.30

27 Underwriters start using statistical scores to help in evaluating loans, learning about new empirical risk attributes, and identifying risks that deserve more emphasis. Interactions between the statistics and manual underwriting’s loan-file experience can help fine tune the power of a scoring system over time, but they also make “natural-state” underwriter versus scoring comparisons more difficult to find and make. Natural experiments are most feasible with an early version of the scoring system, such as in Makuch’s comparison cited above.

28 Freddie Mac quality control underwriters, sampling manually from the pool, warned of risk. The 700-loan sample was scored (using Freddie Mac’s GMW, the FICO score variant), finding in a matter of hours that about half of the loans were investment quality. Review underwriters (third-party and from the lender) then carefully re-underwrote the full 1,000 loan files, taking more than six months, and finding in the end about half of the loans to be investment quality—but not the same group identified by the scoring method (there was some overlap). After quality control resampling to approve the reviewers’ ratings, Freddie Mac obtained recourse on those loans designated by the review underwriters as non-investment quality.

29 I thank Molly Boesel for her assistance on the analysis summarized in figures 2 and 4.

30 On the foreclosure rate, results for the underwriters were sharply the opposite of expectations, but very much as expected for the scores. While it is difficult to verify, and no difference in servicing practice was stipulated, it seems quite possible that the lender serviced and/or absorbed the losses differently on the non-investment quality underwriter rating loans (since these loans had Freddie Mac recourse). Thus, the delinquency rate comparisons may be most fair to the underwriter ratings—still leaving them performing only marginally well at best, and much worse than the scores.
Figure 2. Scores versus Underwriters on Marginal Gray-Area Loans, December 1993 to April 1994 Affordable Loan Pool

Note: Represents a three-year performance of mortgage score risk rankings versus underwriter risk rankings on a sample of more than 700 special affordable loans. Both underwriters and the score cutpoint found approximately half of the loans to be investment quality—although, despite overlap, not the same loans.
Use of the scores alone, in this case, could have significantly reduced Freddie Mac’s uncovered losses and processing costs. The sample limits our ability to draw strong general conclusions, but the results suggest that marginal gray-area loan applications may be quite difficult judgments for manual underwriters. Statistically based customized scores seem to provide a more accurate means of identifying and quantifying the true risks. Without such knowledge, higher uncertainty may incline traditional underwriting to be more conservative or charge higher-risk premiums for marginal borrowers.31

It is worth noting that if figure 2 reflects a general difficulty that traditional manual underwriting has had with marginal loans, it does not imply that this underwriting, on the whole, was not predictive, only that it was less predictive, on the whole, than scores (similar to Makuch’s finding cited above). It also does not imply that scores are perfect or that experienced manual underwriting cannot add insights to scores. Some problems in identifying and incorporating such added insights, however, are that (1) they can be buried within statistical “noise” generated by mistaken manual judgments, and (2) added manual loan reviews add back costs, which may not be justified by the benefits.

Other Traditionally Marginal and Higher-Risk Applicants and Borrowers

Other evidence that the marginal accuracy of traditional underwriting could be improved by AU includes Freddie Mac’s preliminary finding from an industry pool of 2,300 loan applications denied in 1992; as many as 25 percent of these denials would have been accepted under AU (Freddie Mac 1996, 22–23). Looking at affordable loans, Chemical Bank in early 1996 reported its own internal study that compared the decisions of manual underwriters to Freddie Mac’s GMW. They found that, compared with manual underwriting, more loans were approved through the GMW. This aligned with internal results from Freddie Mac’s 1994 analysis of 771 Affordable Gold GMW-pilot loans with GMW scores and manual underwriter decisions provided.

More recently, industry interest has grown in comparing scoring with manual underwriting on traditional subprime mortgages. Preliminary Freddie Mac estimates have suggested that 10 to 35 percent of subprime borrowers could qualify for a prime loan (Freddie Mac 1996, 23–24). The conclusion has been supported by early and developing versions of a customized (proprietary) Freddie Mac model for subprime. Figure 3 illustrates the analysis, using an early-version subprime default scorecard based on a sample of nearly 25,000 loans originated by several subprime lenders (with manual underwriting grades available for about 60 percent).32 The loans had three to five years of default experience, and also were compared with reported manual-grade performance on about 50,000 to 100,000 comparable subprime loans (Bear, Stearns, and Company 1999) and a random sample of more than 230,000 Freddie Mac prime loans.

31 Similar informational inefficiency may also extend into subprime (further discussion follows). The GMW, at the time, also approved somewhat more of these affordable-loan borrowers than the review underwriters—for both minority and nonminority groups. Subsequent comparison to Freddie Mac’s current LP has found these gains in the accept rates to be larger. (As planned, the GMW was phased out as LP expanded.)

32 I thank Ying Li and Andy Jaske for work on the development of this model.
Figure 3. Subprime Marginal Default Rate Comparison:
Manual Underwriting Grades versus Customized Scores

Note: Compares marginal three- to five-year default rates from traditional underwriter-grade risk rankings versus a statistical default scorecard's risk rankings. The sample was 25,000 subprime loans (with supplemental reported grade performance from 50,000 to 100,000 more loans).
Figure 3 suggests that a customized subprime model can outperform manual subprime grades (which are predictive). The model identifies 30 percent of subprime borrowers with lower default rates than that identified by the best manual grade (high \(A^-\)) and another 30 percent with default rates higher than high \(A^-\) but lower than other \(A^-\) grades (other \(A^-\) comprised most of the \(A^-\) loans in these data). Compared with the prime-loan-margin benchmark default rate (normalized to 1.0), approximately 30 percent of subprime borrowers appear to be sustainable within-margin prime-loan candidates using the model. In contrast, even the high \(A^-\) manual-rated loans lie outside the prime-loan margin, implying that it is difficult to identify or qualify eligible subprime borrowers as prime with prevalent manual subprime underwriting methods. The statistical model also identifies 20 percent with higher default rates than identified by the worst manual grade (\(D\)).

The keys to the model's superior performance in identifying likely low and high rates of default appear to be (1) use of more variables than those in a typical subprime underwriting guideline sheet, and (2) statistical optimization in risk-factor tradeoffs. Subprime-loan underwriters must judgmentally classify loans into \(A^-/B/C/D\), because typical underwriting guideline sheets provide univariate knockout rules but not statistical scorecard-like risk tradeoffs or risk-factor weights. Some subprime underwriting guidelines also appear to be adapted from prime-loan underwriting and are less predictive for subprime. This particular early Freddie Mac model result does not account for the uncertainty regarding expected subprime default. Other institutions are increasingly using statistical models and automated tools to evaluate subprime risk.

**Fair Lending, Disparate Impact, and Other Mortgage Scoring Concerns**

Banking regulators and legal analysts have tended to cautiously endorse mortgage scoring models and AU, acknowledging their value but also warning about potential pitfalls (e.g.,

---

33 Subprime data findings made available by the Mortgage Information Corporation (MIC) and others also have shown the manual grades to be generally predictive of default risk. (These grades are relatively quite rules-based, e.g., on credit and debt.) See, for example, Office of Thrift Supervision (2000) for an analysis using the MIC subprime data.

34 Consistent with these findings (1) Equicredit, a subprime lender that is now part of Bank of America, recently reported that its underwriters found fewer than 1 percent of their borrowers eligible for upgrade to prime, and (2) the improvement over subprime grades brought by a scoring model (chart similar to figure 3) also has been shown by Wilson (1999). Wilson’s findings were based on a default scorecard estimated using Transamerica subprime data.

35 Subprime underwriting guidelines commonly classify all applicants with debt ratios above 50, for example, as below \(A^-\), and likewise for all with 60 days past due mortgage payments within 12 months, and so forth. A debt ratio of, for example, 51 with a clean mortgage history (and, e.g., all or most other factors pointing to \(A^-\) or better) is classified traditionally as \(A^-\) or \(B\), according to the underwriter’s judgment.

36 Freddie Mac’s developments continue amid growing signs that AU and other models likely will replace subprime manual underwriting, grading, and pricing systems in the coming years. As occurred previously in the prime market, some pioneering institutions have already been using scoring models for subprime mortgages for some time, including Commercial Credit (now part of Citicorp) and Household (Transamerica implemented its model shortly before exiting subprime mortgages). Also as occurred first in the prime market, mentored–artificial intelligence (rules-based) AU systems have grown as well (e.g., Internet based). Some subprime investors and bond insurers have used a model from MIC, and Bear Stearns uses a model. First Union and Countrywide have announced other recent AU and risk-based pricing models, applicable to prime or subprime. Fair Isaac’s new “Next Generation” generic score is better customized to handle subprime.
Dennis 1995; Inside Mortgage Finance 1995h; OCC 1997). Others have been more skeptical, with positive aspects noted but concerns ranging from fair lending questions to potential fraud, the ability of models to withstand changing economic conditions, privacy issues, and errors and incompleteness in credit data (e.g., Fishbein 1996; Fitch IBCA 1995; Kunkel 1995). Such concerns are not surprising, given the magnitude of the changes that have taken place and those that are still underway. Higher-powered new technologies require higher-powered knowledge and prudent application. To a considerable degree also, many concerns about shortcomings and issues in the previous mortgage finance system have now been retranslated and redirected toward the new increasingly scoring- and model-based system.

One retranslation of an old issue is concern about the ability of scoring models to predict the default risk of low-income applicants and minority applicants. Over time, such concerns have diminished. Freddie Mac, for example, has compiled extensive evidence of the predictive strength of its LP system and key components across demographic groups. Fair Isaac has conducted its own study of its FICO credit scores (Fair, Isaac and Company 1996), with similar findings. Other institutions, including Fannie Mae, have now compiled similar results. Interestingly, analyses such as these have held scoring to a higher statistical standard than most traditional manual underwriting. As evidenced in figures 2 and 3, scoring could well surpass traditional underwriting’s predictive accuracy on marginal and higher-risk low-income and minority borrowers.

Disparate impact and other fair lending questions reflect other concerns. For example, observers have criticized the potential misuse of credit scores (Washington Post 1995, 14–15): “People are looking for ways to make their work simpler. I don’t think the scores or process of scoring is in any way discriminatory. But how it is used can result in discrimination.”

Prevention of misuse in scoring requires strong industry communication, education, and system design. But it also argues against giving too much latitude to subjective judgments that can bring improper interpretations. And it can be analogous to computer phobia—society counters it, but we do not ask computer makers to stop making and improving computers (nor hold them liable) when a hacker infects systems with a virus.

Disparate impact, affecting some minority groups, is the disproportionate effect of higher denial rates (or pricing effects) that any underwriting system likely has—manual or AU-based—and is a consequence of statistical group disparities in wealth, income, and other fac-

37 Part of this is summarized in Freddie Mac (1996, 20–21, 27–28) and Mahoney and Zorn (1996).

38 Identical predictive strength of scoring models across all borrower groups would be a statistical rarity indeed—actually an impossibility. For groups where a model might be (at least marginally) less predictive than for others, at what point does such a difference become material? One surely relevant comparison (where possible) is with traditional manual underwriting’s ability to likewise sort out any differences in key risk determinants across groups. Regulation B of the federal Equal Credit Opportunity Act places certain restrictions on scorecard models (e.g., inability to give extra score credit to single female borrowers, even though this group has been commonly found to be lower risk, all else constant), which has been a topic of discussion in scorecard modeling for many years.

39 In the beginning, Fannie Mae’s former chairman, James Johnson, expressed similar concerns (Inside Mortgage Finance 1996b). Also see Fannie Mae (2000, 7–8).

40 Similar concerns have been expressed about potential disparate treatment of “refer” applicants. One strong way to counter this possibility is to expand the “accepts” by eliminating “refers,” which Freddie Mac’s LP did in 1998.
tors (Fannie Mae 2000, 6–7) affecting the group distributions of mortgage applicants’ collateral, capacity, and credit. The law does not prohibit disparate impact but, rather, requires business justification. Among alternative underwriting systems—manual, artificial intelligence—or rules-based AU, or statistical scoring-based AU—statistical scoring has an inherent advantage of empirically demonstrable business justification, linked most directly and strongly to profit and loss. Scoring also facilitates a more effective and objective search for model/system alternatives that might meet the business need with less disparate impact. These strengths bring more measurable standards but also should not hold scoring to a double standard against alternative underwriting methods and previous practices.

One outgrowth of fair lending concerns, along with more clearly measurable credit quality standards in scoring, has been a growing—and historically overdue—mortgage industry outreach to consumers on understanding credit-record data and its importance to loan evaluations. Another outgrowth has been debate over the openness of AU systems. As this article has been revised for publication, Fannie Mae, Freddie Mac, and Fair Isaac have all recently released publicly the factors used in their DU, LP, and FICO scoring models, giving interested consumers unprecedented new access to the criteria by which their mortgage applications are statistically judged.

For public policy, market expansion and disparate impact in the aggregate translate to scoring’s effect on total and group loan volumes. Volume effects, holding constant the exogenous risk, new-applicant and cyclical factors, and other market factors, are hard to pin down. But various strands of suggestive evidence have been offered in the foregoing text. Freddie Mac (1996), after considering the available evidence, concluded that, as has been seen in other forms of consumer credit, the most likely net impacts of scoring and AU would be greater democratization of credit, with expanded volumes and more favorable pricing for most borrowers. Since then,  

41 It is important to note that most lower-income and minority loan applicants do not have gray-area or higher risk. Moreover, owing to their larger share of the population overall, nonminority borrowers receive the majority of subprime loans, for example. At the same time, it is well known that relatively higher percentages of some lower-income and some minority applicant groups have tended to fall into the more marginal- and higher-risk ranges.

42 AU and electronic data bring efficiencies to self-testing on fair lending generally (Covington 2000), but scoring-based AU also forces institutions to ask if judgmental underwriting aspects applied to loans are truly predictive of risk. Recent interest in fair lending concerns regarding AU has reflected a new and ongoing HUD investigation of the Fannie Mae DU and Freddie Mac LP systems. Interest is not limited to mortgage industry observers, since in the sizable banking and finance universe of scoring overall, mortgage lending is uniquely situated for use (or abuse) of the HMDA data on borrowers’ race. Fair Isaac regularly comments on fair lending matters in its Viewpoints publication. Makuch (1998b, 74–76) discusses fair lending laws and regulations in the development of scoring systems.

43 For earlier discussion, see Inside Mortgage Technology (2000, 18–20). Some have even suggested that all statistical scorecards should be “glass boxes,” fully viewable to all. (Freddie Mac’s GMW headed this way, and the insurance industry in general has had some greater information sharing.) But it is not at all clear that this would promote the proper balance across consumer rights, industry information sharing, and competition. For consumer and lender access to the greatest variety of lowest-priced competitive credit options and innovation, competitive proprietary scorecards and technologies, balanced by more openness, could well be best.

44 By sharpening the risk-identification lens, scoring creates new winners and losers compared with previous industry underwriting procedures. A “swap in” or “swap up” to lower-priced, lower-risk categories for borrowers previously considered too risky for such classification is countered by a “swap out” or “swap down” to higher-priced, higher-risk categories for borrowers who previously slipped through with unnoticed, overlooked, or mistakenly discounted risk characteristics. Net volume impacts depend on which effect is larger—the swap-in/up or the swap-out/down—for the industry overall and for lower-income and minority borrowers. The social gain or cost weights that should be assigned are arguably not identical if some swapped-out/down borrowers are dissuaded from high-risk loans.
minority and low-income loan volumes have continued to rise at much faster rates than white and higher-income volumes.\textsuperscript{45}

As a final piece of suggestive evidence, figure 4 shows the percent of loans for minorities and whites that were granted by known early mortgage scoring adopters versus other institutions (from 1993 to 1995, based on Freddie Mac loan purchases).\textsuperscript{46} The known scoring institutions had both a higher percentage and a faster positive change in minority loans during 1993 to 1995, lending further support to the view that statistical AU may raise lower-income and minority loan volumes in comparison to traditional underwriting. To control for possible exogenous factors, the \textit{all other} loans were reweighted to reflect the LTV, region, and loan-purpose distribution of the known scoring loans. These variables explained most of the minority-lending difference, leaving marginally greater growth in minority lending in the known scoring institutions, although the extent to which the controlled differences (e.g., LTV) were endogenous to factors in the scoring systems in use is unclear.

Other, at least potentially, valid AU concerns noted earlier (e.g., potential fraud) are beyond the scope of this article. The same is true of various econometric questions in scorecard model development—for example, potential omitted-variable bias (Avery et al. 2000)—that may not be adequately dealt with by all models and modelers. At least in principle, most developers and users of scoring tools and AU systems have considerable profit incentives to avoid such pitfalls. This is not intended to pay such concerns short shrift, or to diminish the need for broad-based and knowledgeable expertise, care, and due diligence in statistical AU and costing/pricing model developments and use. Makuch (1998a, 6) estimates that more than 95 percent of credit card applications require no assistance from a human underwriter, implying that a growing percentage of mortgage applicants also will be underwritten and priced in this way as statistical mortgage model developments and uses continue to mature.

Conclusion and a Look Ahead at Likely Developments and Research and Development Needs

This article has reviewed the 1990s shift from traditional manual mortgage underwriting to mortgage scoring models and statistical AU. AU has become the predominant mortgage underwriting method, helping to catalyze a broad wave of technological advancement, with still further significant changes likely ahead. Discussing the history plus underlying factors and concerns in the growth of statistical AU, the article especially considered the question of how far AU should be relied upon to evaluate more marginal and higher-risk applicants (tradi-

\textsuperscript{45} See, for example, \textit{Inside Mortgage Finance} (2000a), which reports on the 1999 HMDA data release. Positive volume effects from statistical AU may exist despite some misperceptions. HMDA conventional-loan denial rates for total and minority applicants also increased for a time after 1995, for example, but this may reflect a mixture of factors, including an increase in applicants with lower down payments and poorer credit histories. Part of this may also have reflected a transition phase, during which underwriters may have made use—or misuse—of both new risk indicators from scores and some increasingly outmoded beliefs in old risk rules and conventions.

\textsuperscript{46} The \textit{known scoring} group included a little more than 65,000, 47,000, and 35,000 loans, respectively, in 1993, 1994, and 1995, while the \textit{all other institutions} group included 2.03 million, 1.19 million, and just over 824,000 loans. The mortgage scoring users for this analysis (from early 1996) may not be an exhaustive list of all such users during 1993 to 1995; however, it does seem clear that, owing to its relative size, the \textit{all other institutions} group was virtually all or heavily made up of institutions that were not making extensive use of mortgage scoring during 1993 to 1995.
Figure 4. Percent Minority and White Mortgage Loans, 1993 through 1995
(Known Mortgage Credit Scoring versus All Other Institutions)

Note: Compares a Freddie Mac sample from 1993 through 1995 of 65,000, 47,000, and 35,000 loans, respectively, from known mortgage scoring institutions versus 2.03 million, 1.19 million, and 824,000 loans from all other institutions (most or all not using scoring). Reweighting controls for LTV, region, and loan purpose. Non-credit scoring institution loans are reweighted using the distribution of the credit scoring institution loans. Variables controlled for are: LTV, region, and loan purpose.
tionally harder cases), showing evidence, and discussing fair lending strengths, that lend support to a cautious, continued expansion of statistical AU into higher-risk loans.

Future growth areas of industry research and development (proprietary and public), education, and resource needs on statistical AU and related areas—in no implied order of importance, nor claim of completeness—seem likely to include:

1. risk- (or value-) based costing and pricing models and systems and AU;
2. fraud models and detection systems in AU;
3. optimizing model/system estimation and delivery techniques and controls;
4. education (industry and consumer) on the factors used, the cost savings, and benefits from data, statistical models, AU systems, and related technologies;
5. economic research on the incidence of AU gains/costs across investors, lenders, and consumers;
6. Internet system growth in AU, risk-based pricing, and so forth;
7. credit-record dynamics and responsive product innovations;
8. further subprime and other market-segment, and product, system customizations;
9. alternative credit histories, thin credit files, data errors, and responsive products;
10. appraisal or home-value estimate streamlining and dynamics and products;
11. economic projections in AU and related systems, and stress tests;
12. modeling questions of potential omitted variables, and general data needs;
13. servicing and behavior scoring models and AU;
14. methods and constraints for maximizing minority loan volumes and other advantages through AU and related statistical decision tools; and
15. borrower retention models, cross-selling, loan streamlining/document reduction, and other lender and borrower AU linkages and customer demands.

For example, while risk- or value-based pricing is not new to mortgages, automated, finer, and more flexible loan-level pricing models/systems, linked with AU models and technologies, seem likely to grow.\(^{47}\) Potentially, all or most loans, regardless of risk, can be quickly approved and offered at statistically appropriate competitive risk-based prices.

\(^{47}\) Such systems must generally model default, prepayment, severity, and capital risks, and so on. See Beidli and Focardi (2000).
In another example, credit-record dynamics research and competition have already led to industry products that reward credit-impaired borrowers making timely mortgage payments with automatic reductions in their interest rates. Similarly, some lenders are already offering automatic rate reductions to most borrowers when prevailing rates fall to make their prepayment option in the money, and AU and related technologies seem likely to facilitate more product innovations aimed at customer attraction and retention (see, e.g., Inside Mortgage Finance 1999a and Harney 2000).

Internet applications of AU have grown substantially, especially in the past two years (most of this in business-to-business broker-lender transactions), with future growth likely in many dimensions. In one of the most intriguing “races” underway, competing systems have sought to bring full AU and pricing and closing of mortgages to a complete transaction, including disclosures, executable in a short space of time over the Internet. In the 1990s, statistical models often led automation (as new models waited for systems and users to advance and adjust), but the decade now begun seems headed toward more of the reverse—new technology, flexibility, and competition placing stronger demands on the timeliness and capabilities of statistical decision tools.

It seems likely that statistical AU for mortgages is not only here to stay but can be expected to keep growing and branching out still further with other models into deeper, often linked, applications in the decade ahead. Realizing and sustaining the full potential of these new technologies will continue to require high levels of institutional energy, experience, care and sensitivity, knowledge and expertise, and business, technological, and modeling sophistication. As in the 1990s, one likely key measure of the strength and resilience of these fundamental changes will be their ability to better democratize and expand mortgage credit, and more affordable mortgage credit generally, to borrowers from all groups.

References


See, for example, Cornwell (1999), Inside Mortgage Finance (1999b), Inside Mortgage Technology (2000), and Strickberger (1999b) for discussions of the ongoing technology races. All-electronic mortgages may or may not use the Internet. A new federal electronic-commerce law that took effect on October 1, 2000 cleared the way for the first of such transactions (Mondor [2000] has an earlier discussion).

Automated appraisals feeding AU systems, and others, for example, have been one of the strongest areas of model demands already for some time (Quinn 2000).


Inside Mortgage Finance. 1999a. October 1, pp. 3–4
KPMG Consulting. 2000. KPMG’s Mortgage Production Performance Study (MorPro). May.


Wilson, Don. 1999. B&C Lending: This Year’s Model. Paper presented at Faulkner & Gray Conference, 12–14 May, La Jolla, CA.
