Using non-traditional data for underwriting loans to thin-file borrowers: Evidence, tips and precautions

Received (in revised form): 29th October, 2007

Michael A. Turner*

is the president and founder of the Political & Economic Research Council (PERC). PERC is a nonprofit public policy research organisation focusing upon issues pertaining to consumer credit, community economic development and economic development. Dr Turner is an international credit reporting policy expert who has worked with industry and governments in more than 12 countries. He currently serves on the Advisory Committee for the Brookings Institution Urban Markets Initiative, and has just completed a two-year term on the US Department of Homeland Security's Data Privacy and Integrity Advisory Committee. Dr Turner earned his PhD in political economy from Columbia University, and his BA in economics from Miami University.

Amita Agarwal

is a director for Chase Credit Card Services, responsible for acquisition risk within the credit card portfolio. Her key responsibility is to maintain and manage all acquisition credit risk strategies across all the channels, products and partners. One of her goals is to continuously explore, test and bring in new opportunities in terms of new updated risk tools to make optimal acquisition risk decisions and create market expansions. Prior to Chase she was at GE Consumer Finance-America, as risk leader for their bankcard product, managing the life cycle risk for the product. Before that she worked for other credit issuers like Fleet Credit Card Services and Advanta. Dr Agarwal received her MA in mathematics in India and her PhD in mathematics/statistics from Syracuse University.

*PERC/Information Policy Institute, 100 Europa Drive, Suite 403, Chapel Hill, NC 27517, USA
Tel: +1 (919) 338 2798 x801; Mobile: +1 (919) 265 3466; e-Fax: +1 (212) 656 1732; Email: turner@infopolicy.org

Abstract Sustainable growth in underserved domestic markets has long been a challenge to lenders. Recent testing with non-traditional data in automated underwriting shows promise as a means to profitably extend credit to the 'thin-file' and 'no-file' populations without assuming undue risk. This area is in its infancy, and is fraught with risk and challenge. Despite the potential, lenders are advised to proceed with caution and should slowly test their way into this segment with the new methods. As this is a slow process, one of the key challenges is to get the needed commitment from the lending institutions. A prudent credit risk officer can harness the power of non-traditional data by taking a disciplined and methodical approach to testing and implementing. This paper demonstrates the value of non-traditional data as a powerful tool for consumer credit risk assessment while highlighting some of the potential risks and precautions that lenders need to think about before using these tools. Special emphasis is placed on paying attention to the capacity of these customers and creating a life cycle strategy for them that includes credit education. This paper presents some empirical test results, and outlines steps that should be taken by lenders to capture the full value of the data while mitigating risk.

Keywords: *credit risk, non-financial payment reporting, thin-file, automated underwriting, alternative data, credit scoring*

THE LIMITS OF AUTOMATED UNDERWRITING: THE THIN-FILE AND NO-FILE POPULATION

Over the past two decades, automated underwriting has become the dominant method for allocating loans for most mainstream lenders in almost every advanced economy. In addition, many consumer lending businesses rely heavily on automated direct mail channels, proactively soliciting worthy prospects for various financial products. These automated systems use an applicant's or prospect's payment history and other behavioural data found in credit files. This enables lenders to abandon the often cruder and more subjective manual credit decisioning methods to match credit offers more precisely to an individual borrower's credit risk, credit capacity and credit worthiness. Relative to manual underwriting, the market's reliance on these statistical credit decisioning methods has been a boon for most borrowers, with the result being increased access to less expensive credit for consumers new to the credit market.[1] Further, consistent automated decisioning based on rules allows systematic audits and easier regulatory compliance for lenders.

However, for those with little or no credit information — the 'thin-file' and 'no-file' populations respectively — neither manual nor automated underwriting offers much by way of a solution to accessing affordable, mainstream credit. The size of this segment has been estimated to range between 35 and 54 million persons in the USA.[2] When confronted with applicants with little or no information, mainstream lenders generally rely on the operating assumption that they are high-risk or that there is not enough information to make a decision to offer them an existing product. In most instances, the result is rejection. Many in this cohort are thus forced to resort to payday lenders, cheque cashing services and subprime lenders. While some in this segment, especially some of those who secure subprime mortgages or a secured product with collateral, do shift to mainstream lending, others remain trapped in a 'credit Catch-22' in which one must have a credit history in order to qualify for credit. Even for those who do get into the financial mainstream via taking up a secure product, the initial transition can be difficult and comes with a cost.

Until recently, the primary obstacle preventing lenders from penetrating this 'final frontier' in mature retail credit markets has been the pronounced lack of systematically collected, standardised, verifiable data upon which to make credit decisions. Now, however, there are a growing number of creative efforts to include non-traditional or non-financial data — payment data from non-financial obligations including utilities, telecoms, and rent — in credit files for use in scoring algorithms.[3]

Credit bureaus, analytics firms and other hybrid players in this emerging niche market have found it a hard sell when making their case to mainstream lenders. This scepticism has been largely attributable to two factors. First is the chicken and egg nature of the issue: credit bureaus had not made the collection of non-traditional data a business imperative owing to the fact that large lenders were not demanding the data. Concomitantly, lenders, for their part, were not demanding and testing the value of the alternate data owing to the fact that credit bureaus had very little data to test. Secondly, there was very little evidence available to convince lenders that the use of non-traditional data for scoring would result in sustainable and profitable lending to the thin-file and no-file populations.

A recent, empirical test of the impacts of including two types of non-traditional data (energy utility and telecoms customer payment data) by the Political & Economic Research Council (PERC) and the Brookings Institution Urban Markets Initiative was conducted.[4] It uses approximately 8 million credit files with full-file (positive and negative) payment information — that is, nearly all the files with such data in TransUnion's national database. Much of the evidence for the present paper is drawn from this study.

THE PROMISE AND LIMITS OF ENERGY UTILITY AND TELECOMS DATA

In its report to Congress on ss. 318 and 319 of the Fair and Accurate Credit Transactions Act 2003, the US Federal Trade Commission concluded that non-traditional data showed some promise as a means of reducing financial exclusion for 54 million Americans, but that it was too early to judge such developmental efforts as were afoot in 2003.[5] The Federal Trade Commission did note some of the cost barriers associated with the collection of rental payment data and regulatory barriers that may preclude the reporting of utility (energy and telephone) customer payment data, but still suggested that utility data appeared to be a very promising candidate for providing information about a borrower's credit worthiness.[6]

For lenders considering whether and how to make use of non-financial data for automated underwriting solutions, the first question that must be answered is which of the data sets that collectively comprise the universe of non-traditional data are likely to be most predictive of consumer credit behaviour? The answer to that question will inevitably vary across credit instruments. For instance, small fixed payments associated with a landline telephone account, taken in isolation, are unlikely to yield much that would be predictive in the context of a home mortgage loan, but may be highly predictive in the context of a credit card offer. As such, an analytical framework for assessing the promise of the varying non-traditional data sets is helpful.

Using the three 'C's to evaluate non-traditional data

In an earlier qualitative assessment of non-traditional data, PERC introduced an analytical framework for assessing the near-term promise of the disparate data sets contained within the universe of non-traditional data.[7] The context for assessing the 'promise' of a given data set was its likely predictiveness of default (being at least 90 days past due), a standard measure for a generic credit scoring model. Absent quantitative measures, PERC ranked non-traditional data sets based upon their value along three dimensions. The criteria are as follows:

- *'Credit-like' vs 'cash-like':* Transactions involving the provision of a good or service in advance of payment exhibited properties more akin to a conventional credit transaction than those requiring payment prior to the provision of a good or service. In addition, transactions occurring repeatedly over time at regular and pre-specified intervals exhibit the greatest 'credit-like' properties. Moreover, providers of these services have an incentive to report in that reporting creates an incentive for consumers to pay on time.

- *Coverage:* In that credit scoring models rely on patterns of behaviour exhibited by large groups of borrowers over time, and given the desire of a lender to extend as much credit as is profitable given a risk threshold, the value of any given non-traditional data set will increase as the percentage of the thin-file and no-file population that it covers increases. Certain services, such as utilities, are far more likely to be widely used.
- *Concentration:* From the perspective of collection, interaction with a relatively low number of data furnishers is more cost-effective. The search costs, contracting costs, data testing and quality control measures, data verification and other transaction costs increase as the number of data furnishers grows.

The matrix in Figure 1 depicts varying non-traditional data sets along these dimensions. Owing to their credit-like properties, the relative concentration of the industries reporting payment data, and the broad coverage of the data sets among the target (thin-file and no-file) populations, energy utility and telecoms customer payment data were gauged as the most promising non-traditional data sets for the near-term.

That these two non-traditional data sets have been identified as potentially the most promising given a specific objective must not be interpreted as a condemnation for the potential usefulness of the other non-traditional data sets. The value of any given non-financial data set is likely to vary across different lines of business (eg auto vs consumer real estate), in different applications (eg bankruptcy model vs new account model), and in different contexts (eg mature economies vs emerging markets). In addition,

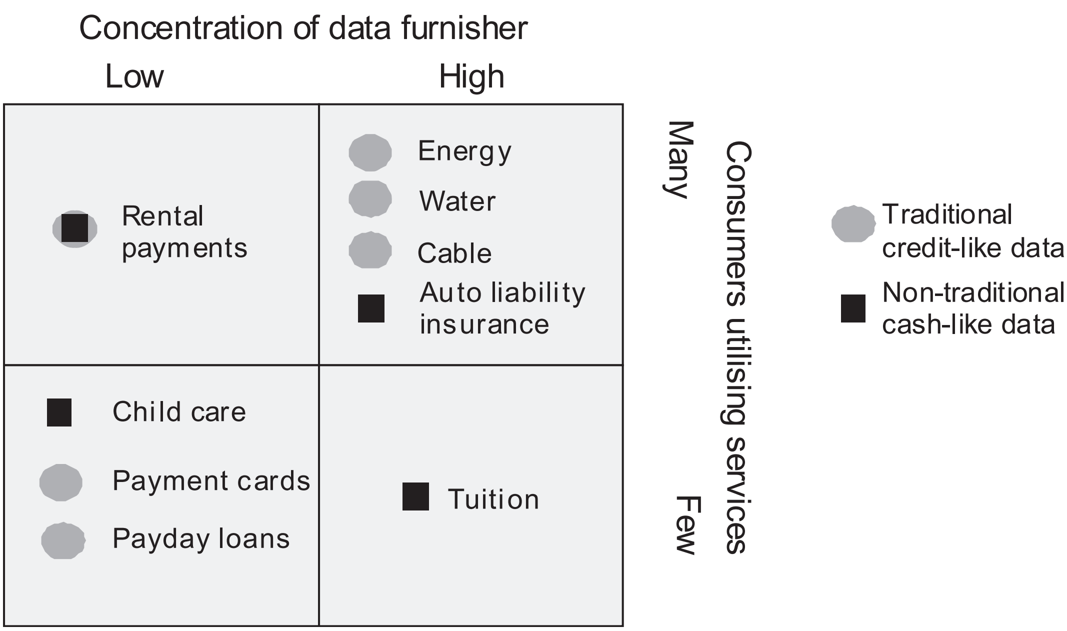

Figure 1: Non-traditional data — concentration, coverage and credit-like-ness

innovative new technologies or applications may be developed that solve some of the business process concerns, such as may be currently happening in the apartment rental data market with the emergence of new software tools that may enable the timely collection of up to 50 per cent of all apartment rental data in the relative near-term.

There are, of course, limits to these sets. Monthly utilities and telecoms are generally a far smaller share of monthly expenditure than, say, rent. As such, lenders may have a peculiarly greater interest in payment histories of rent, auto insurance and tuition. Yet, to the extent that access to credit can ratchet up a consumer's credit file and assist them towards asset building, utilities and telecoms payment data provide considerable practical promise. Moreover, many lenders will require more than one or two trade lines in order to extend some forms of credit. More non-traditional trade lines thus assist greatly.

RESULTS FROM EMPIRICAL TESTS: MEASURING THE PROMISE OF ALTERNATIVE DATA

The empirical analysis of the impact of alternative data used all of the credit files in the TransUnion database that had at least one fully reported telecom or utility payment history, with the payment record going back at least one year, as of March 2005. Due to the lack of such payments being fully reported, this amounted to around 8 million credit files or only 4 per cent of the credit files in their database. Still, this makes for a huge sample, capable of producing extremely statistically significant results.

Using several commercially-used credit scoring/screening models (VantageScore, TransRisk new account, TransRisk bankruptcy, a second bankruptcy model from a large financial institution, and a mortgage screening model from a major lender), credit scores, with and without the alternative payment data, were obtained for these 8 million credit files in March 2005.[8] Following this, the predictiveness of these scores, with and without the alternative data, was then assessed during a year-long observation period (March 2005 through March 2006) over the outcomes each model was designed to predict, such as a serious delinquency or bankruptcy. Therefore, real alternative data, in real credit files, using real credit scores and real outcomes were used. In addition, to benchmark various findings, around 4 million randomly selected files that contained no telecom or utility payment data were used.

Impact of adding utility and/or telecom payments to credit files

As shown in Table 1, 9.6 per cent of the consumers with reported utility payments and 14 per cent of consumers with reported telecom payments have no 'traditional' payments reported in their credit files. As many credit score models can produce a score with only one trade line, including those used in this study, these individuals move from being unscoreable to being scoreable because their energy or telecom payment data are reported. Thus, the use of non-financial data all but eliminates the condition of being unscoreable in these samples. In addition, many of the consumers have multiple utility or telecom payments reported. This explains why the proportion of consumers with no payment histories prior to the inclusion of the alternative

Table 1: Impact of reporting utility and telecom payments on consumer credit files

Total number of payment histories	Consumers with utility payments reported		Consumers with telecom payments reported	
	Including utility payments	Excluding utility payments	Including telecom payments	Excluding telecom payments
Thin file, <3	11.8%	17.0%	18.4%	23.0%
0	–	9.6%	–	14.0%
1	7.7%	4.0%	13.4%	4.9%
2	4.1%	3.4%	5.0%	4.1%
Thick file, ⩾3	88.2%	83.0%	81.6%	77.0%
Sample size	7,519,020	7,519,020	590,795	590,795

payment is greater than the proportion of those with only one payment history after being included.

As depicted in Figure 2, only around 4 per cent of the general population see changes in their scores of greater than 50 points, with about an equal share seeing rises as falls. Notably, nearly 45 per cent see no change in their scores at all. The greatest practical impact appears to be the nearly 10 per cent who go from being unscoreable to scoreable.

While the impact of including the non-financial data seems to wash out when viewing the broader population, the story is dramatically different when viewing the thin-file and no-file population. As much derogatory energy utility and telecoms payment data are already reported to consumer reporting

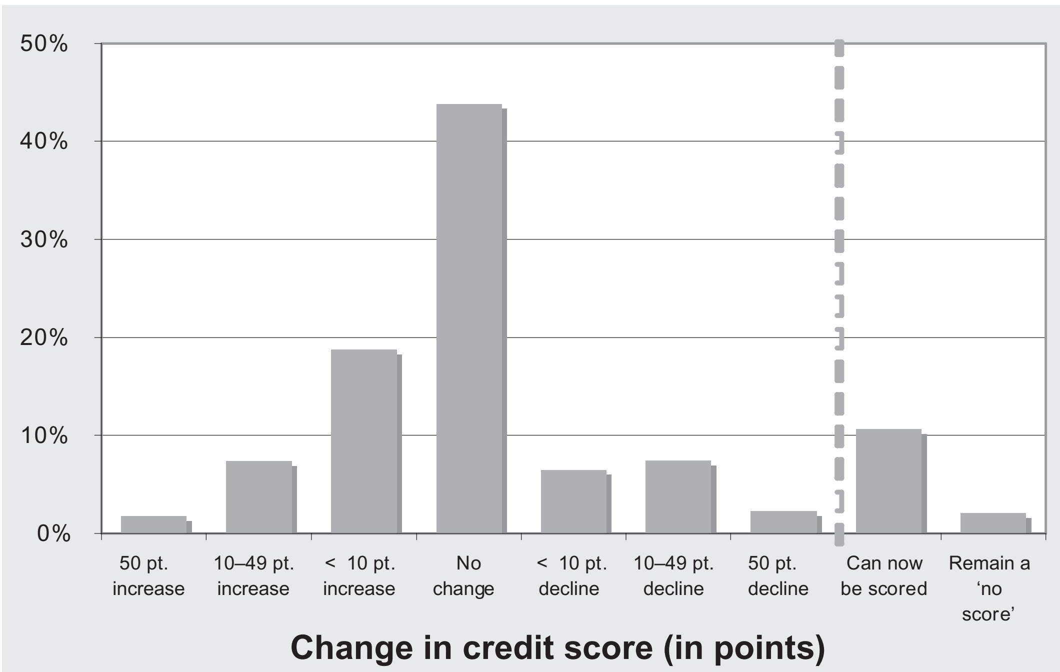

Figure 2: Impact of reporting utility payments on VantageScore

agencies (CRAs) directly or indirectly (through collections agencies), the inclusion of fully-reported payment data — that is, positive and negative payment data — tends to have a net positive impact on the score distribution for thin-file and no-file Americans. Additionally, those with just one or two trade lines see much larger changes in their scores with the addition of another trade line or two, with less than one-quarter of such individuals having no score changes or changes of less than 10 points.

The actual score distributions (for VantageScore) with and without the utility payment data included are shown in Figure 3. The principal shift in the score distribution when utility payment data are added to the scoring model appears to be a shift in mass away from 'unscoreable' to scores between 501 and 800, with little change occurring in the super prime range (scores above 800). The score distribution of the 10 per cent of consumers who can be scored only when utility data are included is also shown and reflects that the bulk of the new entrants obtain scores fairly uniformly distributed in the 501–800 range. Thus, the people brought into the system with the addition of the utility data are not brought in only to be on the bottom rung.

Results for the telecom sample are qualitatively similar to those for the utility sample and available in the full report.[9]

Table 2 reveals the improvement in scoring model performance with the inclusion of utility payment data in the consumer credit files. Estimates are reported for the entire sample with utility payments and for just those with 'thin-files', ie those with fewer than

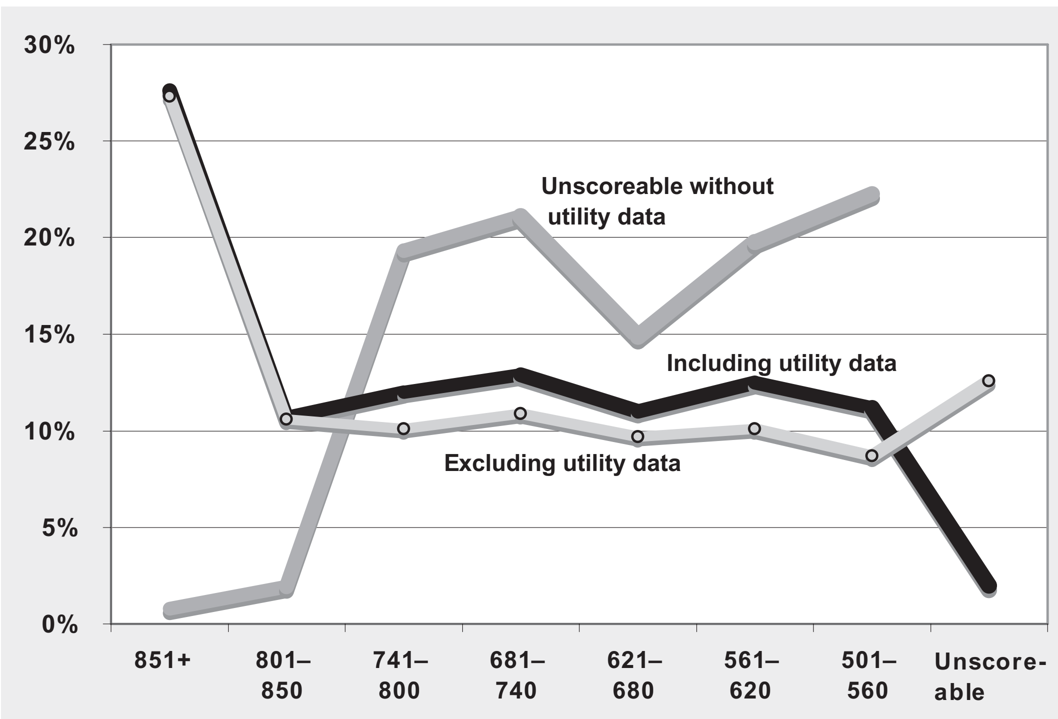

Figure 3: VantageScore distributions

Table 2: Percentage increase in Kolmogorov-Smirnoff test (KS) with inclusion of utility payment data

Model	All consumers Including those with no score (#1)	Excluding those with no score (#2)	Thin-file consumers Including those with no score (#3)	Excluding those with no score (#4)
VantageScore	9.8%	2.2%	329%	7.8%
TransRisk New Account	5.1%	2.5%	293%	6.1%
TransRisk Bankruptcy	13.5%	0.5%	335%	3.5%
Bankruptcy Model II	13.8%	0.8%	359%	5.0%
Sample size	6,211,323	5,439,844	1,280,553	369,903

three traditional payment histories in their credit files. Secondly, calculations were carried out in two ways. The first and third columns include those with no credit scores, which are assumed to be high-risk (putting them at the bottom of the risk ranking). For the results reported in the second and fourth columns, calculations were based only on those consumers who were scoreable with *and* without the utility payment data and, hence, include only those with some traditional payment histories.

As expected, these calculations indicate a much larger improvement in predicting payment outcomes and bankruptcies among those with fewer trade lines. They also indicate that predicting payment outcomes and bankruptcies is greatly improved when using utility data for those with no traditional payment histories compared with just assuming the unscoreable are high-risk. It is also important to emphasise that, in the general sample and among those who are scoreable without the utility data, a modest improvement in the fit of the scoring models is still witnessed with the inclusion of the utility data (second column). This means that the additional information brings new consumers into the mainstream financial system while allowing for a better risk sorting of those already in the system.

It is also important to note that the models in Table 2 were not specifically optimised for utility or telecom payment data (usually just treating them as general trade lines) as such data are not yet widely reported. As such data come online in greater quantities it is almost certain that further improvements in model fit will follow.

In a second test of whether including utility data lends predictiveness or additional predictiveness in assessing payment outcomes, some very basic regression analysis was performed. Using a sample of those with both utility payments and traditional payments reported, two regressions were run. First, whether a consumer had (>90 days past due (DPD)) delinquency on a traditional trade between March 2004 and March 2005 was regressed on whether a consumer had a serious delinquency on any trade the following year, March 2005 to March 2006. The R^2 for this regression was 0.21. Secondly, an additional explanatory variable was added: whether the consumer had a serious delinquency on a utility trade (March 2004 to March 2005). The R^2 for this regression was 0.30, a 40 per cent increase. While this exercise is admittedly crude, it does indicate that the alternative payment data contain additional useful information for predicting future payment outcomes.

For lenders, an improved ability to distinguish good risks from bad risks should translate to lower rates of delinquency (or default) for a given acceptance rate, a greater acceptance rate for a given target delinquency rate, or some mix of the two. This is seen in Table 3 and Figure 4 which are based on actual occurrences of serious delinquencies (> 90 days past due on any account) during the year observation period.

The higher acceptance rates for given delinquency rates and lower delinquency rates for given acceptance rates with the inclusion of the utility data are due to two factors. First, there is an improved ability to distinguish good risks from bad risks, particularly among the thin-file population, those with just one or two traditional trade lines, for whom an

Table 3: Serious delinquencies by acceptance rates: VantageScore model

Acceptance rate (%)	Consumers with utility trades Including utilities (%)	Excluding utilities (%)
30	0.9	1.1
40	1.2	1.5
50	1.8	2.3
60	3.0	4.2
70	5.4	8.1
80	9.5	13.8

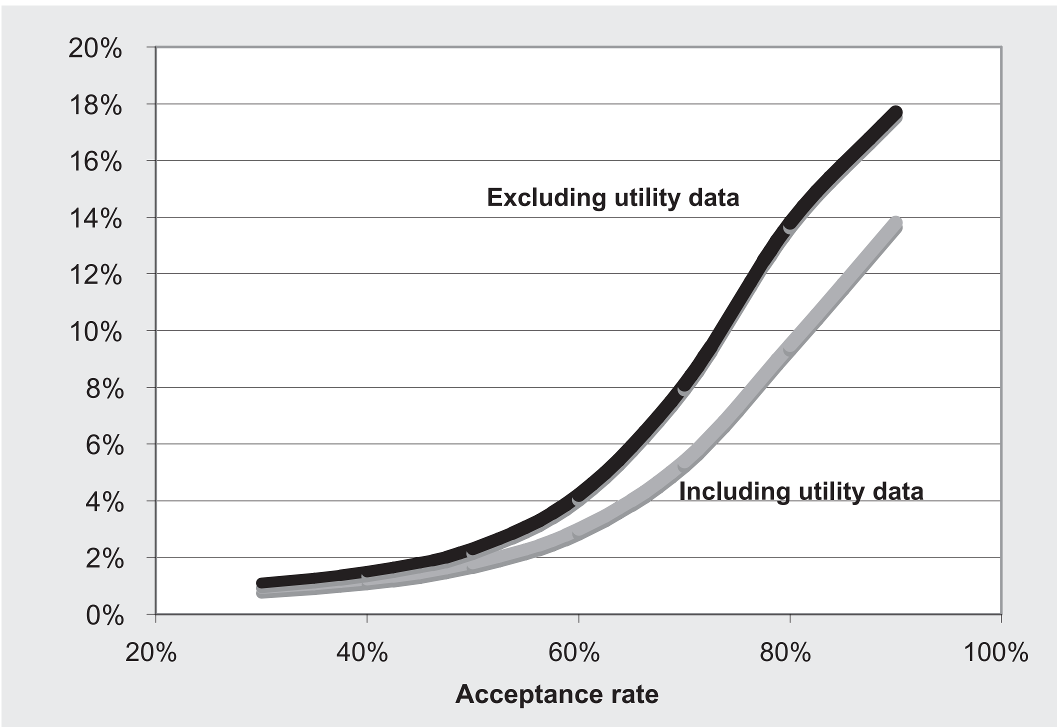

Figure 4: The trade-off between delinquencies and acceptance

additional trade line should have a practical impact. Secondly, with the inclusion of the utility payment data, roughly 10 per cent of the sample becomes scoreable, representing an additional 10 per cent of consumers who can be served.

As seen visually in Figure 4, the number of borrowers accepted can rise by roughly 10 per cent with the inclusion of the utility data without increasing delinquencies (defaults).

A LENDER'S VIEW AND APPROACH

Given today's growth challenges in consumer lending, this 'alternative' information definitely seems to add value in predicting risk and enhancing a lender's ability to underwrite and approve its applicants that were denied credit in the past. Ability to drive initiatives that harness the value of this new information and create viable tools will, no doubt, challenge organisations and affect their future growth potential. However, it is strongly recommended that these initiatives be part of strategic growth plans for all lending organisations.

As the market is becoming mature and saturated, one of the key challenges for the industry is to grow organically. Some key reasons are as follows:

- Attracting customers via automated direct mail channels has become less effective. In reference to the credit card industry alone, total mailing volume for 2006 was 9.2 billion pieces, or about three per week per credit-worthy household. Response rates have fallen from 1.4 per cent in 1995 to 0.3 per cent in 2004.[10] This results in a high cost for acquiring these customers and accounts.
- In the internet era, more and more customers have access to the internet and their shopping behaviour has significantly changed. More prospects are reaching lenders proactively than being solicited through direct mail. This means more thin and no-files are applying for credit. A lender's inability to underwrite in these segments not only affects growth, but also creates a negative consumer experience.
- Many businesses are partnering with the lenders to facilitate the sale of their products. These products could be wide-ranging from instalment loans for home remodelling projects to a US$40 in-store purchase for a retailer. These low-dollar purchases can often be easily underwritten using minimal information about the customer without incurring too much risk, creating the desired experience, both for the customer and business partner.

Lenders are constantly building new and better tools to target segments that they have already been targeting. However, the no-file and thin-file population is a segment in which they have never operated before. The empirical evidence to date suggests non-traditional data add value in differentiating good risk from bad risk and in improving a lender's underwriting capabilities in these segments.

Making the decision to commit to non-traditional data

Every organisation has its own process for investment decisions and prioritisation. One of the challenges is that these initiatives require dedicated resources, time and dollars. Sometimes it is hard to quantify the long-term financial benefits to rationalise these initiatives, gather support to conduct the initial investment, prioritise and execute them.

The market is still evolving from the data perspective. There are many different alternatives available. It is a challenge for any organisation to devote enough resources and to have the discipline to test all the alternatives, understand the value as it relates to its products and target markets, and make the right selection for its business.

Even if the organisation values the available non-traditional data sources for decisioning, given that these tools are in their infancy stage, successfully integrating these data sources into lenders' decisioning platforms remains a challenge. This makes testing, evaluation and comparison more difficult for lenders.

Credit is extended to consumers in the form of various products, mortgages, instalment loans, credit cards, secured cards, etc. Corporations need to test cautiously and learn in any new environment. Extending an unsecured loan for US$50,000 is different from giving customers an unsecured loan for US$3,000. There is a strong recognition that the product features and life cycle management components have to be carefully designed and tested to be successful in this market.

When focusing on a segment that was denied credit in the past due to 'no or little information' it must be recognised that many of these customers are new to credit, and may not be accustomed to handling credit. Initiatives that focus on such a segment should include the development of a life cycle strategy, a strong component of credit education, and a process to monitor performance closely — all of which require additional dedicated resources and strong commitment. Relatively modest investments upstream, however, could stave off considerable losses downstream, providing a successful tool for the business to grow in future.

Testing and comparisons: Why history matters

The first step is to identify the optimal channels in which to test these tools. The optimal tool is the one that has highest penetration, best predictive power to differentiate risk, and maximises the benefit to the lender after taking cost into consideration. The direct mail channel allows a lender to boost the list of potential prospects as well as back-end approval to book additional accounts.

It is important to understand the penetration of these tools in the target segment. Even if the tools differentiate risk well, low penetration within the lender's target segment does not allow lenders to optimise benefits from these tools.

Given the cost of acquiring some of these non-traditional data sets and solutions, it may be desirable to design a strategy that uses these tools effectively, incurring minimal incremental cost and maximising benefit. For example, one may choose to alter the sequencing to deploy a cheaper tool first in the process, reducing the overall cost by using the more expensive tool only on selected applicants.

From some of the testing that has been done, the incremental value of these tools diminishes as more and more traditional data become available. Thus, there may be a value in using these tools only on certain segments of customers.

It is important to have a well-considered test plan to gain maximum understanding. The testing could be done in various ways, as described below.

Retro-scoring

Retro-scoring the files of existing customers with various sources of non-traditional data sets and using their performance to validate tools will allow credit granters to read the results quickly and compare the performances for various tools. This approach has some constraints. It enables lenders to understand performance only in the existing target market, but does not allow any swap-in opportunity for new prospects. In addition, data providers often do not have the capability to retro-score a file.

- *Pros:*
  - quick read;
  - provide comparisons among various tools within the existing parameters.
- *Cons:*
  - data provider capabilities to retro score on sample/performance data of the portfolio;
  - no swap-in opportunity.

Testing in tag mode

Testing in tag mode requires a lender to first build the capability to integrate non-traditional data into its decisioning system. Lenders then accept applications on the margins with various non-traditional data appended to these applications. No decision is made using these new tools. This information, along with the real performance down the line, is used to evaluate the value of each piece of non-traditional data and to provide a comparison among all the tools. Once the winner is identified, this provides a lender with an opportunity to swap in a new population.

- *Pros:*
  - designed to provide swap in opportunities for future;
  - provide comparisons among various tools.
- *Cons:*
  - longer to read results and deploy;
  - build pipe into the decisioning system.

Live testing

Live testing of these tools can be conducted if some preliminary work has narrowed the choices of tools to be tested to one or two. The challenge is still to create a feed for each one of these tools into the decisioning system.

- *Pros:*
  - designed to provide swap in opportunities.
- *Cons:*
  - building feed into the system and corresponding strategies;
  - longer to read results and deploy.

Recommendations for testing: Keeping an eye on the prize

Once the organisation has understood the value of exploring these new data sources, it should keep the following points in mind when testing non-traditional data in credit decisioning tools.

- *Focus on these initiatives:* Clearly communicate goals across functional areas. Ensure that the focus is maintained to implement and monitor/track results. Achieve operational effectiveness across all the functional areas. These initiatives also entail vendor management.
- *Proof of concept before infrastructure development:* This sequence enables rapid prototyping and validation of results. Continually test the processes, including

operational support and procedures.

- *Design a clear test plan:* Create a well-considered test plan to include as many available data sources as possible and to gain maximum learning. Establish success metrics as input to potential roll-out decisions.
- *Cost/benefit analysis:* Integrate these new tools to maximise benefit and minimise cost. Consider the lifetime value of the prospective customer when selecting which tool to use. Some segments may only justify lower cost tools. Further, the right sequencing of the tools to maximise profits is essential. Wherever possible, deploy more expensive tools towards the end of the process only on a targeted segment.
- *Develop life cycle management strategies for this segment:* Lenders should ease these customers into the world of mainstream credit. Start with a product that has little exposure and can help customers to learn to manage credit before opening them to traditional products and strategies.
- *Invest in customer education:* Thin-file and no-file segments are still new to credit and require a credit education and repeat contact strategy to manage the accounts effectively. One method that has demonstrated promise is borrower education administered by the original lender. Lenders willing to invest in user-friendly online and paper content, as well as spend some time discussing credit responsibility — and cautioning the borrower to resist the temptation of accepting more credit than they either need or can handle — may prove to be an effective protection against the pitfalls of bandwagoning.
- *Monitor, monitor, monitor:* Closely monitor the stability of these tools and recalibrate them as these segments become more mainstream in the business and as tools using non-traditional data gain favour among other lenders. The addition of a single traditional trade line in their credit file fundamentally changes how other lenders perceive the borrower. Emerging evidence suggests that the extension of credit based upon non-traditional data to a borrower with little or no credit history is quickly followed by the extension of multiple credit offers from other borrowers. In many instances, the size of the credit on subsequent offers is substantially higher than that offered by the original lender. This can result in over-extension and potential universal default. One large card issuer reported that their borrowers tended to be delinquent or in default on subsequent lines of credit (those extended by lenders based upon the new traditional credit trade line) but continued to perform well on the original account. While the precise formula is unknown, lenders must develop some ability to assess risk and capacity for first-time borrowers with thin credit files. Bandwagoning on a new borrower can result in a suboptimal outcome for subsequent lenders.

The value of a pragmatic implementation approach: Blind faith versus 'slow and grow'

Here it is necessary to insert an element of pragmatism. While early tests of energy utility and telecoms payment data have yielded promising results, lenders must proceed with caution. Moving too fast without properly testing the value of each data set in specific contexts can be disastrous. One major lender reported major losses in its automobile loan silo on loans underwritten using an early variant of FICO Expansion.[11]

In this case, the lender interpreted the correlation between the Expansion score and a generic Beacon score as literal, and began underwriting auto loans using the Expansion score in isolation. The lender then reported that default rates soared in the first quarter, and many loans were ultimately written off. Despite this, the lender remains interested in non-traditional data as an input in credit risk scoring models.

Experiences like those of the lender referenced above have led many to question how a single payment in a small amount can be equated with more conventional credit trade lines. This is a legitimate concern. Fortunately, it can be answered empirically. And while it would be folly to underwrite a home mortgage loan or a large auto loan based primarily upon a single US$25 retail trade payment over 24 months, it would be equally foolish to entirely dismiss its value.

In stark contrast to the lender that took great losses in their auto silo, GE Money has employed a 'low and grow' strategy when using non-traditional data to extend credit to thin-file applicants who usually have no prior credit history. GE Money will offer a borrower a low credit limit at competitive rates, and track their performance over time.

If the borrower meets certain performance criteria, their credit limit will be gradually increased. GE Money's ultimate objective is to deepen and broaden the relationship over time, migrating performing borrowers from a card into depository accounts, auto, student and consumer real estate. Its internal testing concludes that at least 40 per cent of the thin-file/no-file population — 14 to 22 million borrowers — can be profitably banked with existing credit instruments.[12]

STATUTORY/REGULATORY BARRIERS TO REPORTING

In four US states (California, New Jersey, Ohio and Texas) there are rules prohibiting the onward transfer of utility and/or telecom customer data.[13] None of these restrictions are specific to consumer credit reporting, and all relate more to enhancing existing state or federal data privacy protections. Discussions are now ongoing with lawmakers in these states to reconsider these laws in light of evidence that fully reporting customer payment data to CRAs would directly benefit many less fortunate citizens in those states. While some progress is being made, in other states, there is newly introduced legislation specific to consumer credit reporting that would impede the ability of energy utility companies to report to CRAs.[14] Federally, the privacy protections of the Telecommunications Act 1996 have reportedly been interpreted by Verizon as permitting the reporting of negative payment data only. As a result of this interpretation, Verizon discontinued fully reporting to CRAs on over 20 million landline customers.

Owing to the existence of these state and federal statutory barriers, as well as the pervasive regulatory uncertainty voiced by state regulators across the country, there is a clear and compelling need for a preemptive federal policy solution. Lenders interested in extending credit to the thin-file/no-file populations must have access to non-traditional data to underwrite such loans. Such data can only be had in sufficient quantity if the existing statutory barriers are removed, and the regulatory uncertainty is abated. In this sense, regulatory reforms remain a priority government affairs issue for lenders.

CONCLUSION

Results from first-generation empirical tests of non-traditional data in credit risk assessment and automated underwriting are promising. Despite the demonstrated potential, lenders would be wise to proceed with caution both in testing the data and when implementing solutions that utilise non-traditional data inputs. Lenders are encouraged to conduct ongoing tag mode tests with non-traditional data across varying segments (cards, consumer real estate, autos, depository etc) to understand the enterprise-wide applications and global value of different data sets. Particular emphasis should be placed upon understanding the value of the data not only for credit risk, but also credit capacity.

Once tools using non-traditional data have been developed and fielded, lenders should invest in thorough consumer education for new borrowers brought in using alternative scoring solutions. In addition, lenders are cautioned to avoid 'bandwagoning' — whereby a borrower brought into the credit mainstream by one lender using non-traditional data is subsequently targeted by other lenders with further offers of credit, often with larger credit limits than the borrower can handle. Such behaviour is a recipe for disaster. Rather, lenders are encouraged to employ a 'low and grow' strategy for expansion into the thin-file and no-file market. Using this approach, borrowers are initially offered low credit limits and competitive rates, and are rewarded with higher limits and lower rates if they perform over time.

Finally, the promise of non-traditional data in underwriting will never be fully realised unless industry and government act to remove barriers to the reporting of such data. These barriers include statutes and regulations that are not specific to credit reporting, but that effectively prohibit energy utility and telecommunications firms from reporting customer data to credit bureaus. Those forward-thinking lenders must make the collection of non-traditional data in consumer credit reports a business priority. Communicating a desire for more non-traditional data to the credit bureaus, and directing government affairs resources to the collective efforts to remove statutory and regulatory barriers are two steps that lenders could take that would help bring to fruition the promise of non-traditional data. If these steps are taken, lenders will enjoy growth in the retail credit market while reducing financial exclusion for millions of borrowers.

References

1 For a discussion of the impacts of automated underwriting on consumer mortgage lending, credit cards and general purpose credit, see: Turner, M. A. (2003) 'The Fair Credit Reporting Act: Access, Efficiency and Opportunity', The National Chamber Foundation of the United States Chamber of Commerce, Washington, DC.

2 The estimates vary widely because two different proxies were used. Fair Isaac Corporation estimates that of the roughly 215 million non-incarcerated adults in the USA, 22 million have no credit files at all. An additional 32 million have 'thin files', meaning files that do not contain sufficient information to calculate a standard credit score. FICO's estimate (54 million) uses the number of 'unbanked' individuals as measured by the Federal Reserve Bank as a proxy for the unscoreable population. See: Lee, W. A. (2004) 'Eyeing the underbanked: Fair Isaac starts a bureau of its own', *American Banker*, June. Experian

bases its estimate (35 million) upon the number of credit eligible people in the USA, and then subtracts the number of individual credit files in its FCRA regulated database, available at http://goliath.ecnext.com/coms2/gi_0199-4905781/Scoring-in-the-unscored-market.html#abstract (accessed 12 September 2007).

3 For a solid description of the actors in this space, and some of their more recent solutions, see Jacobs, K. (2006) 'Reaching Deeper: Using Alternative Data Sources to Increase the Efficacy of Credit Scoring', Center for Financial Services Innovation, Chicago, IL.

4 Turner, M. A., Lee, A., Schnare, A., Varghese, R. and Walker, P. (2006) 'Give Credit Where Credit Is Due: Increasing Access to Affordable Mainstream Credit Using Alternative Data', The Brookings Institution, Washington, DC.

5 Federal Trade Commission (2004) 'Report to Congress Under Sections 318 and 319 of the Fair and Accurate Credit Transactions Act of 2003', Federal Trade Commission, Washington, DC. Sections 318(a)(2)(D) and (E) of the FACT Act require the Commission to study: 'any common financial transactions that are not generally reported to the consumer reporting agencies, but would provide useful information in determining the credit worthiness of consumers'; and 'any action that might be taken within a voluntary reporting system to encourage the reporting of [these] types of transactions . . .'.

6 *Ibid*., p. 81.

7 Turner, M. A. (2005) 'Giving Underserved Consumers Better Access to the Credit System: The Promise of Non-traditional Data', The Information Policy Institute, New York. See also: Afshar, A. (2005) 'Use of alternative credit data offers promise, raises issues', *New England Community Developments*, Issue 1, Third Quarter, pp. 1–6.

8 Although these consumers are concentrated in a handful of states, due to just a handful of telecom and utility providers fully reporting, there appears to be no reason to believe this population is atypical of the larger telecom and utility customer population. See the full report for basic sociodemographic comparisons of the population with fully reported alternative payment information for a nationally representative sample of consumers with no alternative data reported.

9 'Give credit where credit is due: Increasing access to affordable mainstream credit using alternative data' is available at www.infopolicy.org/publications.htm.

10 Also in 2006, 86 per cent of credit card direct mail included online options. See: Bruene, J. (2007), available at: http://www.netbanker.com/creditdebit_cards/ (accessed 29 October 2007).

11 Interview with risk officer from a major multinational lender at the Credit Scoring and Risk Strategy Association 14th Annual Conference at Taboo Island, Canada, 29th May, 2007.

12 Oliver, W. (2007) Comments delivered to experts' roundtable at Extending Credit: Helping Americans Build Solid Financial Futures, Brookings Institution, Washington, DC, 9th January, available at: http://www.infopolicy.org/pdf/20070109.pdf (accessed 5th February, 2008).

13 Turner, M. A. (2005) 'Giving Underserved Consumers Better Access to the Credit System: The Promise of Non-traditional Data', The Information Policy Institute, New York.

14 For a thorough discussion, see Turner, M. A. (2007) 'The promise of non-financial data: how using energy utility and telecoms payment data can help millions build assets', *PERC Issue Brief*, Vol. 7, No. 1, available at: www.infopolicy.org/publications/ADI%20II%20Policy%20Brief.pdf (accessed 29 October 2007).