Effect of Positive Credit Bureau Data in Credit Risk Models

Michiko I. Wolcott
Equifax Predictive Sciences, Atlanta, GA, U.S.A.
michiko.wolcott@equifax.com

August 2007

Abstract: This paper examines the value of the positive bureau data from the perspective of credit risk scoring. Through several case studies from various industries and various countries, the effect of including and excluding positive credit bureau data in credit risk models is measured. A wide range of situations are analyzed, and examples of the business impact of the positive data are estimated under various scenarios. The results show that, while the majority of the discriminatory power in credit risk models comes from negative information, positive bureau data have a significant impact on the quality of risk decisioning, even with less-than-perfect coverage.

Introduction

The scope of credit bureau data can range from complete account-level information including balance and payment information to negative information only. Whether the bureau includes positive credit data or includes negative only, the coverage of the bureau data varies by country; in addition, the percentage of the population that is economically active varies widely especially in developing countries, making the true value of the positive bureau data difficult to quantify. Nevertheless, many creditors are interested in understanding the value of the positive bureau data, especially where bureaus house negative data only, at the least to see whether sharing such data with other creditors is worth considering.

It is intuitive that the positive credit information can have a substantial impact on the marketing side of the credit business, since with positive credit information, usage and profitability may be predicted more easily. The impact to the credit risk decisioning is perhaps slightly more subtle. It is generally understood by the proponents of positive credit information sharing that sharing positive credit information can have a favorable impact on the credit market, not only from the point of view of the creditor but also from that of the consumer. While the main discriminatory power of consumer creditworthiness comes from the delinquency information, positive information can increase the accuracy of credit risk decisioning.

In some cases the decision to share positive information is controlled by legislation in the respective country. In many other cases, however, it is influenced by the concerns of the creditors over transparency and thus increased competition. Creditors are generally reluctant to share more information with other creditors than absolutely necessary, and positive data sharing is far less appealing to the creditors than negative data sharing, perhaps because there is less to lose by sharing negative information.

In decisioning that uses credit registries housing only negative credit information, the credit registries are viewed primarily as blacklists. The assumption is that by taking adverse actions against consumers who have delinquencies, the riskiest of the consumers would be filtered out. However, this approach is not only suboptimal from the point of view of risk control, but also stunts the growth of the credit market since competitive offers cannot be made to consumers that are highly creditworthy. Powell et al. (2004) further notes that the value of proof of good payment history is greater than the value of coming off a negative blacklist. In addition, the lack of positive information in credit bureaus can deny the opportunity for reestablishing the creditworthiness of the consumers who have had delinquencies in the past—if a consumer with past delinquencies is now making an effort to repair his or her credit history by making all payments on time, this effort cannot be rewarded because the only information available on the credit registry would be the history of delinquencies.

In the end, positive credit data sharing in credit registries contributes to the overall consumer economic growth. From the creditors' point of view, it can improve the risk decisioning and increase profitability. Even the simple fact that shared positive credit data provide additional information to severely limited information available by negative-only credit registry likely leads to more informed decisions. This in turn can have the effect of reducing delinquencies and permitting the creditor to take on more accounts without taking on additional risks. From the consumers' point of view, it can improve access to credit since creditors are conceivably able to make better decisions and therefore can accept more consumers without taking on additional risk. In addition, better risk decisioning results in more competitive offers by the creditors that are more favorable to consumers. Furthermore, Stiglitz and Weiss (1981) and Barron and Staten (2003) explain the idea of adverse selection in which non-competitive pricing limits credit access to the creditworthy consumers and results in higher-risked portfolios.

Statistical studies of the value of positive credit information sharing in credit scoring models in developing markets have focused mostly from the perspective of public credit registries rather than private registries, notably by Powell et al. (2004). As public credit registries are controlled by the central bank or the banking superintendency, focusing on public credit registries leaves out an important part of the population, particularly in countries where the proportion of consumers with a banking relationship is still low and access to banking is limited to the higher socioeconomic classes. In many countries, the economically developing segment of the population relies on credit from retailers, finance companies, telecommunications and utility service providers, etc., and this segment represents a large and important part of the credit market. Furthermore, the consumer behavior on these non-bank accounts tends to be distinct from bank trades.

This study examines the range of the impact of the positive credit information sharing in private credit registries (which can also include the data distributed by the public credit registry of the respective country) through four case studies from varying markets with varying levels of coverage. Specifically, this study measures the impact of the positive credit information in risk models built using bureau data and therefore on strategic business decisions. In each case, the credit bureau data are available at the account level for both positive and negative information, aggregated into a standard modeling attribute set which allows the calibration of the impact by type of the account, type of the creditor, etc. This is important, since detailed account-level information is not available in some public credit registries that contain positive information, losing the flexibility and therefore reducing the predictiveness of the data. The four cases selected for this study are: a custom credit card application risk model (United States), a generic risk model (Honduras), a custom wireless thin-file application risk model (United States), and a generic acquisition risk model (Argentina).

Methodology

The basic methodology of this study follows that of Barren and Staten (2003) and Powell et al. (2004). Two models are constructed for each case: one that considers all positive and negative bureau data available as candidate independent variables and another that considers only negative bureau data as candidate independent variables to simulate a negative-only registry. Any attribute that requires the use of positive bureau information is excluded from consideration in the negative-only model. The results are then compared and the impact to the decisioning is measured from the business perspective.

The same sample and performance definition were used for both models in each case, and methodologies and modeling approaches were held as constant as possible between the two models to insure a valid comparison. Additional care was taken so that the results would be as realistic as possible. First, all appropriate legal restrictions for each country were applied (data access, account exclusions, etc.). Second, since inquiry attributes are generally available in the bureaus regardless of the availability of positive information, they were considered in all models. Third, any business requirements within the context of each case, particularly from the cultural perspective, were included.

Logistic regression was used for model estimation in each case. The dependent variable was binary, taking on a value of 1 if bad according to the respective definition, 0 otherwise. The bad definitions used in this analysis focused on severe delinquencies, generally 90 days past due or worse during the performance window. While the dependent variables were defined over specific trades in some cases, only bureau data were considered as independent attributes. Where necessary, bias correction methods were applied as appropriate for the case so that the analysis sample would closely represent the through-the-door population. The same final sample was used to build and assess the two models for each case, so that the only difference would be the available set of candidate independent variables.

The results were measured in three ways: over the total population, over the segment of the population without delinquencies ("clean"), and over the segment of the population with current or past delinquency ("dirty"). The first measures the overall impact of the positive information on credit risk models using bureau data. In the situation where decisioning is based on negative data only, the clean segment represents the consumers that are typically treated in the same way by the creditors because there is little additional information about their creditworthiness, and the dirty segment represents the consumers that are typically declined or heavily screened out. (For the purpose of defining the dirty segment, any past-due account is considered delinquent, since the severity of the delinquency is not always readily apparent in some negative only databases.)

Three tables are presented in each case. The first contains the overall sample statistics, showing the bad rate and the segment breakdown. This sets the general stage for the case. The second shows the cumulative capture rate of bads by percentile, measuring the ability of the models to isolate the bads in the worst percentiles of the population. This equates to the proportion of the bads to be eliminated given a reject rate. The third contains the cumulative bad rate by percentile, showing the portfolio bad rate for a given approval rate. All statistics and results presented below are on a hold-out sample, after the fit of the models were confirmed by comparing the results over the development and the hold-out samples.

Case 1. Credit Card Application Risk Model, United States

The population of interest in this case was all consumer applicants for a credit card with a regional bank in United States. Applications from May 2002 to December 2002 were matched to archived credit databases at the time of application, and the performance was observed for 24 months following the application date. Bad was defined as 90 days past due or worse on the account during the performance window. The depth of the credit file ranged widely in this sample; however, consumers with bankruptcy on file as well as those who were considered credit-dormant were excluded according to the credit policy of the bank.

In the United States, the coverage of positive credit information is assumed to be almost 100% for practical purposes. The shared positive information available includes payment history (generally indefinite amount of history with some exceptions), credit limit or original loan amount, type of account, type of creditor (i.e., industry), and some information related to collaterals where applicable. The motive or the intended use of the borrowed fund can often be identified through type of creditor and/or various narratives.

As can be seen in Table 1a, the difference in the bad rate between the clean and the dirty segments is not as large as one might expect. This immediately suggests that negative-information-only risk decisioning based on a blacklist-type approach may be inadequate for this population. In addition, approximately two-thirds of the population has some history of delinquency, and decisioning that relies strictly on negative data may lead to some kind of adverse action on two-thirds of the population in this case.

Table 1a. Analysis Sample Statistics, Case 1

	Count	%	Bad	Bad Rate
Total	*19,840*	*100%*	*1,942*	*10%*
Clean	6,992	35%	595	9%
Dirty	12,848	65%	1,347	10%

Table 1b shows the relative separation power of the two models. While the model with negative data is effective in isolating the bads in the worst part of the population, including positive information in the model greatly improves the performance of the model, even in the dirty segment. This improvement is particularly substantial in the clean segment, where the only credit information that can be used to assess credit risk would be inquiries if not for the positive information.

Table 1b. Capture Rate of Bads, Case 1

Segment	Model	%Bads Captured in the Worst:				
		10%	20%	30%	40%	50%
Total	With Positive Data	34%	52%	63%	73%	81%
	Lift[1] vs. Negative Only	*63%*	*47%*	*36%*	*30%*	*24%*
	Negative Data Only	21%	36%	47%	56%	66%
Clean	With Positive Data	30%	44%	57%	66%	77%
	Lift vs. Negative Only	*100%*	*80%*	*55%*	*35%*	*28%*
	Negative Data Only	15%	25%	37%	49%	60%
Dirty	With Positive Data	36%	53%	66%	75%	83%
	Lift vs. Negative Only	*63%*	*43%*	*30%*	*25%*	*21%*
	Negative Data Only	22%	37%	51%	60%	69%

Table 1c shows the bad rate for various approval rate scenarios. While the model with negative information only is still effective in identifying the less risky consumers, it is evident that the model with positive information improves this identification substantially. Using the approval rate for the population at the time of the observation of approximately 40%, the approved population would have a bad rate of 6.0% using the model with the negative data only. In comparison, maintaining the same approval rate, using the model with positive data would result in a bad rate of 3.2%, cutting the portfolio bad rate almost in half. On the other hand, if the creditor wishes to maintain the bad rate of 6.0%, the model with positive data would approve more than 80%, effectively doubling the approval volume. Risk management can become more efficient with positive data: the

[1] Lift is calculated as the improvement in percentage; specifically,

$$\frac{\%\text{bads captured by the model with positive data}}{\%\text{bads captured by the negative only model}} - 1$$

creditor can realize a substantial reduction in the portfolio bad rate and/or a substantial increase in the portfolio volume without taking on additional risks.

The scenario is very similar between the clean segment and the dirty segment, save for the top 10% of the population. The reduction in bad rate due to the inclusion of the positive information is similar in the respective best percentiles. This suggests that, with nearly perfect coverage, using positive credit information in a risk model helps identify the best applicants regardless of the past payment history.

Table 1c. Bad Rate by Approval Rate, Case 1

Segment	Model	Cumulative Bad Rate in the Top:							
		10%	20%	30%	40%	50%	60%	70%	80%
Total	With Positive Data	2.9%	2.6%	2.8%	3.2%	3.6%	4.4%	5.1%	5.8%
	Reduction[2] vs. Neg. Only	*34%*	*43%*	*46%*	*47%*	*45%*	*38%*	*31%*	*26%*
	Negative Data Only	4.4%	4.6%	5.3%	6.0%	6.7%	7.2%	7.5%	7.9%
Clean	With Positive Data	2.6%	2.7%	2.8%	3.2%	3.9%	4.8%	5.2%	5.9%
	Reduction vs. Neg. Only	*53%*	*38%*	*45%*	*40%*	*43%*	*34%*	*32%*	*26%*
	Negative Data Only	5.4%	4.4%	5.1%	5.4%	6.8%	7.2%	7.7%	8.0%
Dirty	With Positive Data	2.8%	2.5%	2.8%	3.1%	3.6%	4.4%	5.1%	6.1%
	Reduction vs. Neg. Only	*23%*	*48%*	*49%*	*48%*	*45%*	*38%*	*31%*	*26%*
	Negative Data Only	3.7%	4.7%	5.6%	6.1%	6.6%	7.0%	7.4%	8.2%

In practice, this has two implications. First, the best consumers with no history of delinquency and low probability of delinquency in the future can be offered a more attractive product to improve the creditor's market position. Second, the consumers with history of delinquency who have since improved their creditworthiness through making payments on time can have better access to credit.

In summary, the following are observed: (i) history of delinquency, while it plays an important role in predicting future delinquency, is not the only thing that predicts future delinquency; and (ii) with a nearly perfect coverage of positive data, identification of both the best and the worst consumers is drastically improved.

[2] Reduction in bad rate is calculated as the relative reduction in percentage; specifically, for a given approval rate (i.e., in the top X% of the population):

$$1 - \frac{\textit{Bad rate using the model with positive data}}{\textit{Bad rate using the negative only model}}$$

<u>Case 2.</u> Generic Risk Model, Honduras

The population of interest in this case was all economically active and credit-active consumers in the Honduran market. A random sample of the consumers in the credit bureau was obtained as of February 2003, and the performance was observed for the following 12 months. Bad was defined as 90 days past due or worse over the bureau data. The bad rate for the sample was 23%, and unlike in the previous case, the difference in bad rate between the clean segment and the dirty segment was substantial as can be seen in Table 2a.

The reporting of positive data for the institutions in the banking system is mandated by the banking superintendency. However, in general, only these institutions can legally access the banking system data, positive or negative. Of the rest of the credit market, credit card and other non-banking/non-finance institutions may report positive information to the bureau per the exchange agreement. For practical purposes, the coverage of the positive data in general for the economically active consumer population may be estimated at approximately 50%. For this analysis, in order to assess the impact of all positive information available in the bureau, including the positive information about the accounts in the banking system, the information access scenario of a banking institution is assumed.

The types of positive information available in the Honduran credit bureau are similar to those of the United States and include payment history (generally most recent six months but can contain some indefinite history), balance, type of account, motive or intended destination of the borrowed funds, age of the account, credit limit or original loan amount, type of the creditor, and some collateral information where applicable.

Table 2a. Analysis Sample Statistics, Case 2

	Count	%	Bad	Bad Rate
Total	*94,690*	*100%*	*21,934*	*23%*
Clean	55,011	58%	3,073	6%
Dirty	39,679	42%	18,861	48%

Under less than perfect coverage of the positive information, the capture rates of the bads over the total population do not appear to vary much between the negative-only model and the model with positive information, as can be seen in Table 2b. This is consistent with the idea that the main predictive power of the risk model comes first from delinquency information. In the dirty segment, there is virtually no difference between the two models in terms of this capture rate. However, the model with positive information substantially outperforms the negative-only model in the clean segment. This implies that, with partial coverage, the first improvement that the positive data provide in credit risk modeling is in assessing the creditworthiness of the consumers with no history of delinquency.

Table 2b. Bads Capture Rate, Case 2

Segment	Model	%Bads Captured in the Worst:				
		10%	20%	30%	40%	50%
Total	With Positive Data	43%	71%	82%	87%	91%
	Lift vs. Negative Only	*0%*	*1%*	*2%*	*3%*	*2%*
	Negative Data Only	43%	70%	80%	85%	90%
Clean	With Positive Data	28%	44%	59%	73%	85%
	Lift vs. Negative Only	*55%*	*23%*	*16%*	*17%*	*18%*
	Negative Data Only	18%	36%	51%	62%	72%
Dirty	With Positive Data	21%	42%	61%	75%	84%
	Lift vs. Negative Only	*0%*	*0%*	*1%*	*1%*	*0%*
	Negative Data Only	21%	42%	61%	74%	84%

This is evident also in Table 2c. The inclusion of positive data here clearly has a bigger impact in assessing the creditworthiness of consumers with no reported history of delinquency. However, the best consumers within the dirty segment are also more easily identified with the model with positive data, particularly in the top 10% to 30% range. Again, the model with positive data provides more insight into consumers whose payment behavior has improved since the reported delinquency.

Table 2c. Bad Rate by Approval Rate, Case 2

Segment	Model	Cumulative Bad Rate in the Top:							
		10%	20%	30%	40%	50%	60%	70%	80%
Total	With Positive Data	1.0%	1.3%	1.8%	3.2%	4.2%	5.0%	6.1%	8.4%
	Reduction vs. Neg. Only	*67%*	*59%*	*44%*	*20%*	*14%*	*14%*	*7%*	*2%*
	Negative Data Only	2.9%	3.1%	3.2%	3.9%	4.8%	5.8%	6.5%	8.6%
Clean	With Positive Data	1.0%	1.0%	1.1%	1.3%	1.7%	2.5%	3.3%	3.9%
	Reduction vs. Neg. Only	*64%*	*67%*	*60%*	*61%*	*45%*	*29%*	*16%*	*13%*
	Negative Data Only	2.7%	2.9%	2.9%	3.4%	3.1%	3.5%	3.9%	4.5%
Dirty	With Positive Data	8.5%	8.3%	9.7%	12.4%	15.3%	20.0%	26.2%	34.6%
	Reduction vs. Neg. Only	*28%*	*20%*	*12%*	*0%*	*1%*	*3%*	*1%*	*0%*
	Negative Data Only	11.9%	10.3%	11.1%	12.4%	15.5%	20.6%	26.4%	34.5%

Since the clean segment is approximately 60% of the population, we consider the scenario in which the approval rate is 60%. The population approved at 60% approval rate using the negative-only model results in a bad rate of 5.8%. Use of the model with positive data results in a relative reduction of the bad rate by 14%, or increase in the approval rate to nearly 70% if the bad rate is to be kept constant. Therefore, while the overall improvement may not be as dramatic as in the case of perfect coverage of the positive data, even partial coverage can improve the risk decisioning, resulting in a more efficient and profitable business.

The results of this case point to two key findings: (i) the positive credit information substantially improves the power of the credit risk models even with less than perfect coverage; and (ii) this improvement first manifests itself in identifying the consumers with better risk, with or without history of delinquency. From the consumer's point of view, this can translate to improved offers and access to credit, contributing to the growth of the credit market.

Case 3. Wireless Thin File Application Risk Model, United States

The population of interest for this case was all consumer applications for a wireless service with a national service provider in United States, whose associated credit files are considered to have relatively little information ("thin files"). "Thin" was defined as having one (1) to three (3) trades on the credit file according to the credit policy of the service provider. Applications from November 2004 and March 2005 were matched to the archived credit data at the time of application, and the performance on the account was observed for eight (8) months following the application date. Bad was defined as involuntary deactivation of the account due to non-payment. As mentioned in Case 1, the positive data coverage in the United States is practically 100%; however, in the telecommunications industry the reporting of the positive account behavior is not as prevalent as in other sectors and service providers must rely heavily on positive information reported for other types of accounts and creditors.

Consumers with inquiry-only files or public-records only files were excluded according to the credit policy of the service provider. This credit policy requires that risk decisions be made based on some proof of the past payment history, good or bad. Interestingly enough, the identification of true inquiry-only or public-records only files would be impossible without the positive information, since in a negative-only scenario there would be no way to know that an account was ever booked if the consumer was never delinquent. In this sense, the consumer would be denied credit if something bad existed on the credit file but cannot be rewarded if he or she had a perfect payment history. The thin-file population in the telecommunications industry is an important yet a sometimes problematic segment of the population, since these accounts are often one of the first accounts with payment obligations that a consumer may obtain. Therefore, having positive information in the credit bureau becomes critical for those consumers that are starting to build a credit profile.

In this case, no consumers are declined outright; rather, the high-risk consumers are required a deposit to open an account. While this makes intuitive sense, ineffective risk assessment can easily result in loss of business, since creditworthy consumers whose risk is overestimated can easily seek more competitive offers from other service providers with better risk assessment.

Table 3a. Analysis Sample Statistics, Case 3

	Count	%	Bad	Bad Rate
Total	*216,517*	*100%*	*56,029*	*26%*
Clean	141,920	66%	30,141	21%
Dirty	74,597	34%	25,888	35%

As can be seen in the table above, the population has an overall bad rate of 26%, and the difference in the risk level between the clean and dirty segments is evident by the difference in the bad rates, albeit not as dramatic as in the Honduras case. The clean segment is about two-thirds of this population, much larger than in the credit card population in Case 1. However, this is partly due to the fact that for consumers to be in the dirty segment in this case, they would have to be delinquent on the first accounts they would ever book. On the other hand, those that are delinquent almost as soon as they establish a credit history may be deemed intuitively as the most risky.

The model with positive information isolates the bads better than the negative-only model, and as may be expected the improvement due to the inclusion of the positive information in the model is more dramatic in the clean segment than in the dirty segment, particularly in the worst 30% of the respective segment. Table 3b shows the details.

Table 3b. Bads Capture Rate, Case 3

Segment	Model	%Bads Captured in the Worst:				
		10%	20%	30%	40%	50%
Total	With Positive Data	17%	32%	46%	58%	69%
	Lift vs. Negative Only	*13%*	*10%*	*9%*	*7%*	*5%*
	Negative Data Only	15%	29%	42%	54%	66%
Clean	With Positive Data	20%	36%	49%	61%	71%
	Lift vs. Negative Only	*18%*	*14%*	*12%*	*10%*	*8%*
	Negative Data Only	17%	31%	44%	55%	65%
Dirty	With Positive Data	13%	25%	36%	47%	57%
	Lift vs. Negative Only	*7%*	*10%*	*9%*	*9%*	*7%*
	Negative Data Only	12%	22%	33%	43%	54%

Consistent with the previous cases, the reduction in risk level of the approved or "no-deposit" population is substantial; at a 50% no-deposit rate, the expected relative reduction in bad rate is about 10%. More realistically, the service provider would aim to accept more consumers without requiring a deposit to improve its competitive position. Therefore, maintaining the bad rate of approximately 18% (17.8%) at the 50% no-deposit rate using the negative-only model, it can accept almost an additional 10% without deposit if it uses the model with the positive information instead.

Table 3c. Bad Rate by Approval Rate, Case 3

Segment	Model	Cumulative Bad Rate in the Top:							
		10%	20%	30%	40%	50%	60%	70%	80%
Total	With Positive Data	7.2%	9.9%	12.0%	13.9%	16.0%	18.2%	20.1%	22.0%
	Reduction vs. Neg. Only	*46%*	*26%*	*16%*	*13%*	*10%*	*8%*	*6%*	*4%*
	Negative Data Only	13.4%	13.5%	14.4%	16.0%	17.8%	19.7%	21.5%	23.0%
Clean	With Positive Data	6.2%	8.0%	9.7%	11.1%	12.5%	13.9%	15.4%	17.1%
	Reduction vs. Neg. Only	*55%*	*41%*	*28%*	*22%*	*15%*	*12%*	*10%*	*6%*
	Negative Data Only	13.7%	13.4%	13.5%	14.2%	14.7%	15.9%	17.0%	18.3%
Dirty	With Positive Data	21.5%	25.0%	27.0%	28.5%	29.6%	30.5%	31.6%	32.7%
	Reduction vs. Neg. Only	*27%*	*16%*	*13%*	*11%*	*8%*	*7%*	*5%*	*3%*
	Negative Data Only	29.4%	29.9%	31.1%	32.0%	32.2%	32.8%	33.1%	33.7%

In the situation where consumers have only a few accounts on which they have payment obligations, having only the delinquency information severely limits the availability of data with which creditors can make risk decisions. In this sense, positive credit data provide additional information for the part of the population where having additional information is critical.

Even in a thin-file population, the inclusion of positive data in the credit risk model has a significant business impact: it can allow the service provider to make better offers without taking on additional risk, potentially improving its competitive position in the marketplace.

Case 4. Generic Acquisition Risk Model, Argentina

In this case, the population of interest was all economically active consumers in Argentina who were seeking additional credit. All consumers with inquiries in January 2006 were selected for this analysis, with a performance window of 12 months following the inquiry date. Bad was defined as 90 days past due or worse on any account reported to the bureau during the performance window. In Argentina creditors may report both positive and negative information or only negative information per the bureau agreement, and the overall coverage of the positive data may be considered greater than the case in Honduras but not a near-perfect coverage as in the United States. The types of positive information available in the bureau are similar to both United States and Honduras.

The bad rate for the total population in this case is 13%, and there is a substantial difference in bad rate between the clean and the dirty segments. Approximately 30% of this population has a history of delinquency; approximately 70% has no delinquency on file.

Table 4a. Analysis Sample Statistics, Case 4

	Count	%	Bad	Bad Rate
Total	*20,919*	*100%*	*2,628*	*13%*
Clean	14,556	70%	922	6%
Dirty	6,363	30%	1,706	27%

The impact of the positive data on the model is again similar to the previous cases. Over the total population, the model with positive data isolates 97% of the bads in the worst 30% of the population, whereas the negative-only model isolates only 87% in the worst 30%. The negative-only model is quite adequate in this case, again consistent with the idea that the main predictive power of the risk model comes first from delinquency information. However, the inclusion of the positive data still provides a substantial improvement in model performance, particularly for the portion of the population without prior history of delinquency.

Table 4b. Bads Capture Rate, Case 4

Segment	Model	%Bads Captured in the Worst: 10%	20%	30%	40%	50%
Total	With Positive Data	69%	96%	97%	97%	97%
	Lift vs. Negative Only	*5%*	*18%*	*12%*	*8%*	*6%*
	Negative Data Only	65%	81%	87%	89%	92%
Clean	With Positive Data	93%	94%	94%	95%	96%
	Lift vs. Negative Only	*38%*	*22%*	*16%*	*11%*	*9%*
	Negative Data Only	68%	77%	81%	86%	89%
Dirty	With Positive Data	34%	67%	94%	97%	98%
	Lift vs. Negative Only	*1%*	*6%*	*17%*	*11%*	*7%*
	Negative Data Only	33%	63%	80%	87%	91%

The results in terms of bad rate in the approved population are shown in Table 4c. At the extreme best portion of the population, the most improvement due to inclusion of the positive information in the model is seen in the clean segment; however, in general the effect is similar with or without prior history of delinquency. The reduction of bad rates is substantial throughout the score ranges and across segments. For example, at a 40% approval rate, the negative-only model produces a bad rate of 1.9%; with the model with positive data, the same approval rate produces a bad rate of 0.5%, almost one-fourth of that with the negative-only model. On the other hand, keeping the bad rate of 1.9%, including positive information in the model would allow a creditor to approve more than 80%, more than doubling the approval volume.

In addition, the identification of the best consumers is improved substantially by inclusion of the positive data in the model. Again, this translates to the creditor's ability to potentially improve offers to the consumers who deserve them, which can enable it to be more competitive.

Table 4c. Bad Rate by Approval Rate, Case 4

Segment	Model	Cumulative Bad Rate in the Top:							
		10%	20%	30%	40%	50%	60%	70%	80%
Total	With Positive Data	0.4%	0.4%	0.4%	0.5%	0.6%	0.6%	0.6%	0.6%
	Reduction vs. Neg. Only	*63%*	*75%*	*77%*	*72%*	*69%*	*71%*	*75%*	*78%*
	Negative Data Only	1.1%	1.6%	1.8%	1.9%	2.0%	2.2%	2.4%	2.9%
Clean	With Positive Data	0.1%	0.2%	0.2%	0.3%	0.5%	0.5%	0.5%	0.5%
	Reduction vs. Neg. Only	*86%*	*79%*	*80%*	*73%*	*69%*	*65%*	*69%*	*74%*
	Negative Data Only	1.0%	1.1%	1.1%	1.3%	1.4%	1.5%	1.7%	1.8%
Dirty	With Positive Data	1.3%	1.0%	1.0%	1.1%	1.2%	1.4%	2.5%	11.0%
	Reduction vs. Neg. Only	*53%*	*75%*	*73%*	*74%*	*75%*	*75%*	*68%*	*11%*
	Negative Data Only	2.7%	4.1%	3.9%	4.3%	4.7%	5.6%	7.7%	12.4%

The conclusions of this case are somewhat similar to Case 1, with a better-than-average coverage of the positive data over the economically active population of consumers. The implications to business are again both in terms of loss reduction as well as in terms of business volume.

Conclusions

The four cases presented above cover a wide range of business scenarios, populations of interest, and positive bureau data coverage. In each case, the results show that the inclusion of positive bureau information in credit risk models can improve the performance of the model substantially. In particular, the following can be concluded:

- When there is near-perfect coverage of positive bureau information, credit risk models improve in the identification of consumers with the best risks as well as those with the worst risks;
- When there is less-than-perfect coverage of positive bureau information, the first impact of the positive information on credit risk models is the improvement in identifying the consumers with the best risks, before the improvement in the identification of the consumers with the worst risks;
- The impact of the positive bureau information in credit risk models is generally greater on consumers without a history of delinquency;
- The positive bureau information acts as additional information in credit risk models for consumers without well-established credit history, improving the performance of the model over this emerging credit population.

The business impacts of the above are likely to be seen in substantially reduced delinquency rates and increased business without taking on additional risks. The increased business can translate to better access to credit from the consumer's point of view, which in turn helps the economic development particularly in the emerging markets. Furthermore, the inclusion of positive bureau data in credit risk models can allow creditors and service providers to make better and more competitive offers, further contributing to the growth of the business and the economy.

When these improvements are translated into amount of revenue and profit, the business case for sharing positive credit information is easy to establish. In addition, the consumers benefit heavily, albeit indirectly, from shared positive credit information, contributing to the general development of the economy.

REFERENCES

Barron, John M., and Michael Staten (2003). "The Value of Comprehensive Credit Reports: Lessons from the U.S. Experience." In Margaret Miller (ed.), *Credit Reporting Systems and the International Economy* (pp.273-310). Cambridge, MA, and London: The MIT Press.

Powell, Andrew, Nataliya Mylenko, Margaret Miller, and Giovanni Majnoni (2004). "Improving Credit Information, Bank Regulation and Supervision: On the Role and Design of Public Credit Registries." World Bank Policy Research Working Paper 3443.

Stiglitz, Joseph E., and Andrew Weiss (1981). "Credit Rationing in Markets with Imperfect Information." *American Economic Review*, 71, 393-410.