Theory and Empirical Evidence for Price-Driven Adverse Selection in Consumer Lending

Robert Phillips[*] Robin Raffard [†]

July 18, 2009

1 Abstract

The purpose of this paper is to describe our planned program of research into investigating and quantifying the phenomenon of *price-driven adverse selection* in retail credit markets. Specifically, we will use historical data on consumer characteristics and loan performance for loans funded at different APR's to estimate the extent to which changing the price (APR) of a loan changes the default rate. We will use this information to develop models that can be used by lenders to better estimate the magnitude of adverse selection in their markets and therefore to determine how to best incorporate adverse selection in their pricing and underwriting decisions.

2 Background and Previous Research

Adverse selection has long been recognized as an important characteristic of credit markets. While adverse selection can take many forms, we consider *price-driven adverse selection* – the phenomenon that, all else being equal, *raising the price of credit charged to a market segment will result in deteriorating average loss behavior for that segment.* The obverse is also true: *lowering the price charged to a segment will result in improved average loss behavior from that segment.* The change in loss behavior may manifest itself in changing probability of default, changing loss-given-default, or both. Price-driven adverse selection is a widely acknowledged phenomenon at commercial lenders, yet, in our experience, many lenders are unable to quantify its effect. An improved ability to quantify price-driven adverse selection would support better credit pricing and underwriting decisions.

The underlying reason for price-driven adverse selection is that customers with a lower probability of default (better credit quality) are, on the average, more price sensitive than customers with a

[*]Columbia University Business School and Nomis Solutions
[†]Nomis Solutions

higher probability of default. As the price is raised, customers with a low probability of default tend to seek out alternatives more quickly than those with higher probability of default – and vice versa. One consequence of price-driven adverse selection is that there are some customer segments that it is not profitable to serve at any price and, as a result, in the words of Stiglitz and Weiss (1985), credit is "rationed" – that is, there are some borrowers who will not be extended credit at any price. In practical terms, lenders typically choose an "underwriting threshold" or cutoff risk and will not accept any customers whose risk exceeds the cutoff. Typically, these thresholds do not depend upon the rate being charged.

Adverse selection has been the topic of hundreds of academic papers since its initial identification in the classic paper of Akerlof (1970). Much of this rich literature has focused on modeling and detecting the existence of adverse selection and moral hazard in insurance markets and markets for various "unique goods" such as thoroughbred racehorses (Wimmer and Chezum [2006]) and used cars (Genesove [1993]). While this research has provided considerable insight, many of the results do not directly translate from insurance and unique goods to consumer credit markets.

There have been a handful of studies that look at adverse selection in consumer credit markets. Edelberg (2004) used data from the triennial Survey of Consumer Finances conducted by the US Department of the Treasury to determine the existence of adverse selection and moral hazard in mortgages and automobile loans. Her conclusions were that there is strong evidence for adverse selection in both types of credit, with weaker evidence of moral hazard. A number of researchers have found strong evidence for price-driven adverse selection in credit-card markets. Ausubel (1999) observed that customers choosing an inferior credit card product *ex ante* exhibited a higher default rate *ex post*. Using response and risk data from a credit card company, Agarwal et. al. (n.d.) found that "consumers who respond to inferior offer types (e.g., higher APR) exhibit worse credit risk characteristic than those responding to superior offer types" (pg. 3) – i.e. there was evidence of adverse selection. Karlan and Zinman (2005) worked with a major South African lender to design a randomized experiment in which 58,000 direct mail offers for credit were randomized in terms of interest rate. Rather surprisingly, they found strong evidence of adverse selection among women and moral hazard among men but not the other way around.

We extend previous research in this area by developing predictive models of adverse selection that can be used by lenders to better estimate the effects of their pricing on the riskiness of their portfolios and that can be used to make better pricing and underwriting decisions. We will use data from different lenders for different products (auto, home-equity lending, and unsecured lending) and across different countries to gain a deeper understanding of the underlying determinants of the phenomenon.

3 Basic Models of Adverse Selection

Adverse selection in lending occurs whenever offering a less-desirable (e.g. higher price) product to the same group of prospective borrowers leads to higher loss rates. The typical explanation proposed for adverse selection is "asymmetric information". In terms of consumer lending, this would mean that the borrower has private "adverse information" not possessed by the lender – information that, if he possessed it, would increase the lender's *ex ante* probability that the borrower would default. The borrower's private information is "above and beyond" the information that the lender extracted from the borrower either through its own efforts (e.g. via a loan application) or from third parties such as the credit bureaus. An example of private adverse information might be a series of bad job reviews that has led the borrower to believe that his job is in jeopardy. This information, which is unknown to the lender, would tend to both increase the probability that this borrower will take the loan at a higher price and that the borrower will default relative to other customers who appear *ex ante* identical to the lender.

While asymmetric information is a plausible explanation, it is not the only possible explanation and may not even be the most reason for price-driven adverse selection. Another possibility is that borrowers who otherwise appear identical might differ in their ability to understand and manage their financial commitments. "Financially savvy" borrowers might simply be better at calculating their expected future ability to repay given the prospects that they face and therefore make better borrowing decisions. Less-savvy borrowers might be less good at estimating their future ability to repay and therefore do not make good borrowing decisions. The less-savvy would thus be more likely to accept higher rates and more likely to default due to exogenous shocks or future financial mismanagement even though they appear *ex ante* identical to the lender.

Another possible explanation for price-driven adverse selection, particularly in sub-prime markets, is the so-called "capacity effect". For most borrowers, repaying loan obligations is a lower priority than paying for food, rent, and taxes. The difference between a borrower's monthly disposable income and her expenditure on food, rent, and other necessities is called the borrower's *capacity*. If the monthly payment for a prospective loan is a large fraction of a borrower's capacity, then an increase in the monthly payment due to a higher APR could increase the probability of default due to unanticipated future expenditures such as a medical expense. Some, but by no means all, lenders estimate a borrower's capacity as part of the underwriting process.

From the point of view of our research, the underlying explanations for price-driven adverse selection are not important. Instead, we accept that it is a real phenomenon and we want to i) estimate its magnitude, ii) understand how its magnitude varies with observable consumer and lending product characteristics and iii) develop models that enable lenders to incorporate the phenomenon into their underwriting and pricing decisions. To do this, we begin with the simple assumption that, *ceteris paribus*, the price-elasticity of borrowers is negatively correlated with their probability of default. If this assumption holds, then a population of borrowers will demonstrate price-driven adverse selection.

Consider a population of "risky" borrowers – that is, a set of borrowers with an expected probability of default greater than zero. A lender has no way to distinguish between the "goods" – those who will repay, and the "bads" – those who won't. It is easy to show that, under fairly simple assumptions, a sufficient condition for price-driven adverse selection to exist in such a population is that price-elasticity is higher among the goods than the bads. For any given price, p, let the demand from goods be $d_g(p)$, the demand from bads be $d_b(p)$ and the total demand be denoted by $D(p) = d_b(p) + d_g(p)$. We can define the loss rate as:

$$LR(p) = \frac{d_b(p)}{D(p)}.$$

Then, we note that:

$$\begin{aligned} LR'(p) &= \frac{d_b'(p)D(p) - D'(p)d_b(p)}{D(p)^2} \\ &= \frac{d_b'(p)d_g(p) - d_g'(p)d_b(p)}{D(p)^2} \\ &= (e_g(p) - e_b(p)) \left[\frac{d_b(p)d_g(p)}{D(p)^2 p} \right], \end{aligned} \tag{1}$$

where $e_i(p)$ is the own-price response elasticity for customers of type $i = b, g$ defined as:

$$e_i(p) = \left| \frac{d_i'(p)p}{d_i(p)} \right|.$$

Since the second term in Equation 1 is always positive, the sign of $LR'(p)$ will depend upon the relative values of $e_g(p)$ and $e_b(p)$. In general, $e_g(p) > e_b(p)$ – that is low-risk (good) customers are more price elastic than high-risk customers, so $LR'(p) > 0$. This means that the loss-rate will increase with increasing price, which is adverse selection. We can state it as a general rule that, *adverse selection in a population will occur whenever high-risk customers are less price elastic than low-risk customers.*

This model implies that the magnitude of price-driven adverse selection should be greater in sub-prime (high-risk) populations than in prime populations. To see why, we can rewrite Equation 1 as:

$$LR'(p) = (e_g(p) - e_b(p))LR(p)(1 - LR(p))/p. \tag{2}$$

A prediction of this simple model is that price-driven adverse selection would be more severe in sub-prime markets than in prime or near-prime populations. This is illustrated in Figure 3 which plots the derivative of the loss rate against the loss rate using the relationship in Equation 2. From the figure, it is clear that if a population consists entirely of “goods” ($LR(p) = 0$) or entirely of “bads” ($LR(p) = 1$), there is no adverse selection. For any value of p, $LR'(p)$ is maximized when $LR(p) = .5$. Since the expected bad rate is typically well below 50% even for highly sub-prime populations, adverse selection increases with the expected risk of a portfolio – specifically, adverse selection is a much stronger effect in sub-prime portfolios than in prime portfolios. In high prime portfolios, in which $LR(p) \approx 0$, the effect is usually negligible. However, in segments with high fractions of risky customers, adverse selection can be very significant.

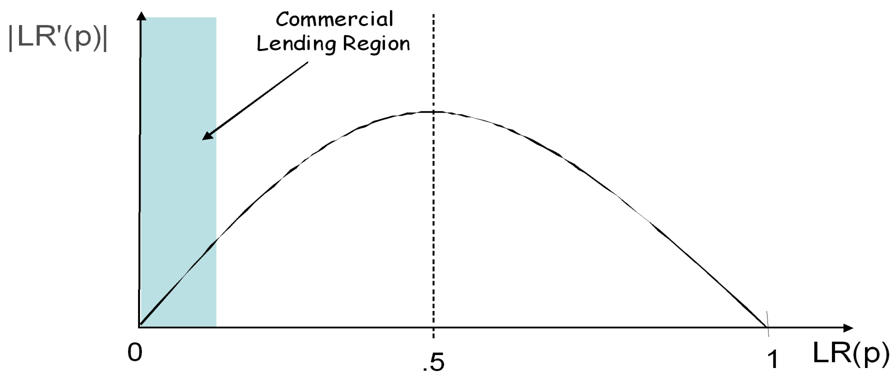

Figure 2. Price-driven adverse selection measured as the derivative of the loss rate with respect to price as a function of the loss rate for the simple model in the text.

This simple model is highly stylized, however it reflects the basic idea underlying our research. In particular, we expect that the expected loss-rate will be better modeled as a continuous function of price across a continuum of customers. However, even the simple model gives some insight into the anticipated relationships between price elasticity and adverse selection. In particular, we hypothesize that the magnitude of price-driven adverse selection will be larger for populations with higher expected loss rates.

4 Estimating Price-Driven Adverse Selection

Most commercial lenders in the US, Canada, and the UK employ some form of risk-based pricing – that is, they charge higher rates to higher-risk borrowers. However, lenders do not vary prices continuously with risk. Rather, they tend to classify prospective lenders into a relatively small numbers of risk categories and (all else being equal) offer the same rate to every borrower within that class (Edelberg, 2006). The classification is typically based on an externally provided credit score (FICO score) supplemented with additional information available to the lender. For example, an on-line auto lender may classify all prospective borrowers into four risk categories. All buyers within the same category applying for the same loan would be quoted the same rate. An individual lender offering a single type of loan might categorize all borrowers with FICO scores between 680 and 720 as "B customers" and quote them an identical rate (APR) – say 7%. Price-driven adverse selection predicts that the fraction of approved applicants who accept a loan at this rate would decrease across the spectrum of credit scores from 680 to 720. Furthermore, if the lender unilaterally raised his APR to 8%, we would expect to the total conversion rate to decrease. Price-driven adverse selection predicts that the percentage decrease in conversion rate would be greater for higher credit score customers than lower within this range.

We call this phenomenon *direct adverse selection*. It is "direct" in the sense that it relates to directly observable customer characteristics (FICO score) and is immediately observable. An example of direct adverse selection from small-business lending is shown in Figure 4. In both Figure 4 A and 4 B, the horizontal axis is the lender-estimated credit score. Higher scores correspond to lower risk. Figure 4 A shows the fraction of approved applications that booked for each score while Figure 4 B shows the linearized average rate offered to borrowers as a function of credit score. It is clear that, even though higher credit quality borrowers were quoted lower rates on the average, they also converted at a lower rate. This is strong evidence of direct adverse selection in this market.

However, price-driven adverse selection also manifests itself in another way. In particular, increasing the rate to borrowers who appear *ex ante* identical in terms of credit quality will *ex post* result in a increase in the average default rate of the borrowers who accept the loan. That is, the expected default rate from a borrower with a credit score of 600 will be greater for a loan at 7% than a loan at 6%. We term this effect *indirect adverse selection*. Unlike direct adverse selection, which is immediately observable, indirect adverse selection requires observation of borrower behavior for some period of time after funding. In particular, repayment behavior needs to be tracked for some period (typically one year or more) to establish statistically meaningful differences in default behavior.

5 Research Plan

The project has access to a number of data sets from commercial lenders in the United States, Canada and the United Kingdom that will be used as part of this project. The data required to estimate *indirect adverse selection* is, for all funded loans over some period:

1. Customer data such as credit score, relationship with lender, etc.

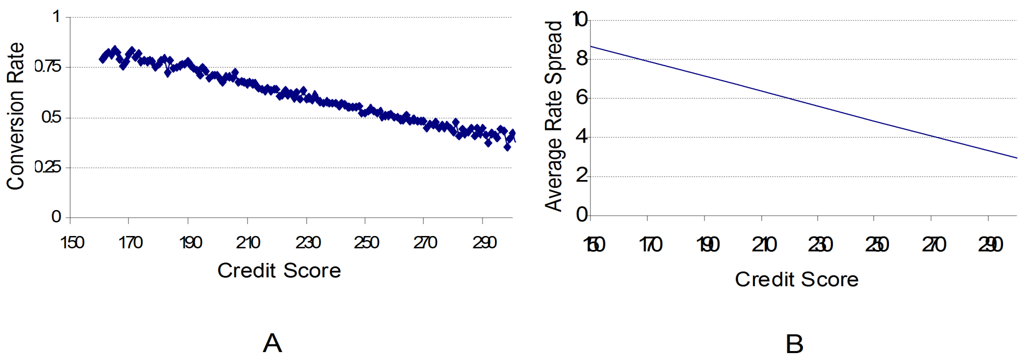

Figure 1: Conversion rate and average rate spread as a function of credit score for two years of small business loan approvals at a major US retail bank. *Credit score* is an internal measure of default risk for an application with higher scores representing lower risk. **A** shows the fraction of approved applications that booked as a function of credit score. **B** shows the linearized ($r^2 = .72$) rate spread (APR - cost-of-funds) for the same population of approved applications.

2. Loan data such as amount, line vs. loan, term, etc.

3. Rate (APR) charged,

4. Payment and default history for some period ≥ 12 months.

We have five data sets that will be used to estimate indirect adverse selection. The characteristics of these data sets are shown in Table 1.

	Line of Business	Country	Date Range	Default Monitoring Period (Months)	Funded Loans	Default Fraction
1.	Personal Lines of Credit	Canada	9/2003 - 8/2005	36	73,002	.55%
2.	Auto, Marine and RV Loans	Canada	8/2004 - 10/2005	36	51,198	1.28%
3.	Unsecured Personal Loans	U.K.	6/2006 - 1/2007	12	82,875	2.05%
4.	Fixed Rate Personal Loans	Canada	11/2002 - 8/2006	24	88,915	1.98%
5.	Variable Rate Personal Loans	Canada	10/2002 - 8/2006	24	164,261	4.05%

Table 1: Sources of lender data to be used in indirect adverse selection estimation. "Default Monitoring Period" is the period for which the performance of each funded loan was tracked. "Default Fraction" is the percentage of funded loans that defaulted over the Default Monitoring Period.

Our estimation of *direct adverse selection* requires a different set of data. Specifically, for each approval loan application, we require:

1. Customer data such as credit score, relationship with lender, etc.

2. Loan data such as amount, line vs. loan, term, etc.

3. Rate (APR) quoted,

4. Outcome: funded versus declined by customer.

We have five data sources of lender data that we are using to estimate direct adverse selection as shown in Table 2.

	Line of Business	Country	Date Range	Approved Applications	Funded Loans	Funding Rate
1.	Auto Loans	Canada	2/2004 - 3/2009	340,952	228,836	67.1%
2.	Personal Lines	Canada	1/2008 - 6/2008	31,064	19,961	64.3%
3.	Personal Loans	U.K.	10/2008 - 2/2009	38,273	28,388	74.2%
4.	On-line Auto Loans	U.S.	7/2002 - 9/2004	321,029	72,535	22.6%
5.	Auto Loans	U.S.	8/2007 - 2/2009	912,098	514,087	56.4 %

Table 2: Sources of lender data to be used in direct adverse selection estimation. "Approved Applications" is the total number of applications approved by the lender during this period. "Funded Loans" is the total number of the approved applications accepted by the customer. The "Funding Rate" is the fraction of Approved Applications that resulted in a Funded Loan.

Estimating both direct and indirect price-driven adverse selection requires price variation. Fortunately, most of the data sets available display significant variation in offering price. Some of this variation is due to price-tests performed by the lenders. In the other cases it is due to discretion available to "front-line lenders" to vary prices to customers as part of a negotiated process. We will exploit this variation to the extent possible in estimating the magnitude of indirect adverse selection for these lenders.

Our research is proceeding in four steps:

1. Cleanse and prepare the data sets.

2. Fit various models of both direct and indirect price-driven adverse selection to the data available.

3. Use the results of step 2 to gain insights into the magnitude and drivers of adverse selection in various settings. That is, to what extent does country (UK versus US versus Canada) make a difference? Direct or indirect channel? Type of loan (auto versus unsecured).

4. Determine the implications for underwriting, pricing, and portfolio risk measurement

6 Acknowledgment

This research is funded in part by the US Federal Deposit Insurance Corporation Center for Financial Research.

References

[1] Agarwal, S., Chomsisengphet, S., and Liu, C. n.d. "Adverse Selection in the Credit Card Market: Evidence from a Natural Experiment" Working Paper, The Federal Reserve Bank of Chicago, available at http://www.geocities.com/ushakri/CreditCardAdvSelection.pdf+agarwal+%22natural+experiment%22.

[2] Akerlof, G. A. 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism". *Quarterly Journal of Economics* **84** 3: 488500.

[3] Ausubel, L. 1999. "Adverse Selection in the Credit Card Market." Working Paper. University of Maryland.

[4] Genesove, D. 1993. "Adverse Selection in the Wholesale Used Car Market," *Journal of Political Economy* **101**: 644-645.

[5] Edelberg, W. 2004. "Testing for Adverse Selection and Moral Hazard in Consumer Loan Markets", Finance and Economics Discussion Series 2004-9, The US Federal Reserve: Washington, DC.

[6] Edelberg, W. 2006. "Risk-based Pricing of Interest Rates for Consumer Loans," *Journal of Monetary Economics* **53** 8: 2283-2298.

[7] Karlan, D. S. and Zinman, J. 2005. "Observing Unobservables: Identifying Information Asymmetries with a Consumer Credit Field Experiment," *Yale University Economic Growth Center Discussion Paper No. 911.* Available at http://ssrn.com/abstract=725563.

[8] Phillips, R.L. 2005. *Pricing and Revenue Optimization.* Stanford University Press: Stanford, CA.

[9] Phillips, R. L. 2009. "An Introduction to Price Optimization for Consumer Credit," forthcoming in *Credit Control.*

[10] Stiglitz, J. E. and Weiss, A., 1981. "Credit Rationing in Markets with Imperfect Information," *American Economic Review*, **71** 3: 393 - 410.

[11] Wimmer, B. S. and Chezum, B., 2006. "Adverse Selection, Seller Effort, and Selection Bias," *Southern Economic Journal*, **73** 1: 201-218.