

A Portfolio View of Consumer Credit*

David K. Musto
Finance Department
The Wharton School
University of Pennsylvania

Nicholas S. Souleles
Finance Department
The Wharton School
University of Pennsylvania
and NBER

September, 2005

Abstract

This paper takes a portfolio view of consumer credit. Default models (credit-risk scores) estimate the probability of default of individual loans. But to compute risk-adjusted returns, lenders also need to know the covariances of the returns on their loans with aggregate returns. Covariances are independently relevant for lenders who care directly about the volatility of their portfolios, e.g. because of Value-at-Risk considerations or the structure of the securitization market. Cross-sectional differences in these covariances also provide insight into the nature of the shocks hitting different types of consumers.

We use a unique panel dataset of credit bureau records to measure the ‘covariance risk’ of individual consumers, i.e., the covariance of their default risk with aggregate consumer default rates, and more generally to analyze the cross-sectional distribution of credit, including the effects of credit scores. We obtain two key sets of results. First, there is significant systematic heterogeneity in covariance risk across consumers with different characteristics. Consumers with high covariance risk tend to also have low credit scores (high default probabilities). Second, the amount of credit obtained by consumers significantly increases with their credit scores, and significantly decreases with their covariance risk (especially revolving credit), though the effect of covariance risk is smaller in magnitude. It appears that some lenders take covariance risk into account, at least in part, in determining the amount of credit they provide.

Keywords: credit supply, consumer credit; default risk, credit scores, loan portfolio analysis

JEL classification: E21, E51, G21

* We would like to thank Anna-Maria Lusardi and seminar participants at Columbia, the Carnegie-Rochester Conference on Public Policy (for which this paper was prepared), the Philadelphia Federal Reserve Bank and various workshops at the University of Chicago and The Wharton School of the University of Pennsylvania. We also thank Lily Fang, Sanjay Chugh, and Nicola Tosini for excellent research assistance. Souleles acknowledges financial support from the Rodney L. White Center for Financial Research, through the NYSE and Merrill Lynch Research Fellowships. We are grateful to the Wharton Financial Institutions Center and Experian for assistance in acquiring the data. All remaining errors are our own.

A basic principle of financial economics is that risk is properly viewed in a portfolio context. That is, the relevant measure of an investment's risk (i.e., the measure relevant for discounting) is the risk that it adds to the portfolio of all investments. The traditional subject of portfolio-risk research has been equities. By contrast consumer loans have received little academic analysis, especially portfolio-level analysis, even though they are typically held in large portfolios of thousands or even millions of loans, orders of magnitude more populated than most equity portfolios. The consumer credit market in the US is quite large, amounting to about \$7 trillion by 2000 (of which about \$5.4 trillion is mortgage debt, \$800 billion is other installment debt, and \$700 billion is revolving debt) [Federal Reserve System (2001)]. Much of this debt is originated by large, national lenders. The debt is typically either held by the originator in large portfolios, or sold into large securitization pools [e.g., see Gorton and Souleles (2004)].

The goal of this paper is to provide a portfolio view of consumer credit. Lenders usually gauge the risk of their consumer loans by credit-risk scores, which measure a consumer's expected probability of default in isolation (a first moment). But the aggregate volatility of a portfolio of loans also depends on the cross-sectional covariances of the default risks of its component loans (second moments), which likely vary across borrowers. For example, the economies of some regions of the country are more cyclical than others. Suppose that default rates in Alaska have a relatively low covariance with default rates in other states; i.e., that Alaskans default relatively less when people in other states default relatively more, and vice versa. Then loans to Alaskans would entail less 'covariance risk' and so bring diversification benefits to a large portfolio of debt, contributing relatively little to the volatility of the portfolio.

This paper uses a unique panel dataset of credit bureau data to construct measures of covariance risk analogous to those from equity research, and more generally to study the cross-

sectional distribution of credit. We analyze the underlying demographic determinants of consumers' covariance risk, and test whether consumers with lower covariance risk (among other characteristics) receive more credit, *ceteris paribus* (*c.p.*). We also measure the effect of the credit scores on the amount of credit, and contrast this with the effect of covariance risk.

There are several reasons for lenders to care about covariance risk. First, the insight of Sharpe (1964) and Lintner (1965) applies to all assets, not just equities: In equilibrium, the market price of a consumer loan should depend on the risk that the loan adds to the market portfolio. If, analogously to the literature on equity investors, we take consumer lenders to be optimizing primarily over their own asset class, in this case consumer loans, then the risk-adjusted expected return of a consumer loan would decline as its covariance with the return on the market portfolio of all consumer loans declines.

Second, there are several important institutional features of the consumer credit market that directly motivate lenders to regulate the volatility of their portfolios. Lenders typically assess the adequacy of their capital through 'Value-at-Risk' calculations, which estimate the probability that losses at the portfolio-level exceed some given threshold [e.g., see Saunders (1999)]. Indeed, the new Basel Accord requires lenders to set their capital levels as a function of such calculations. Also, many lenders rely on securitization to finance a large part of their loans. If total losses in a pool of securitized loans exceed a given threshold, the securitization typically either defaults or enters early amortization, a pre-payment risk for the lender. The portfolio risk of the pool determines how the securitization must be structured to avoid such outcomes and receive desired bond ratings.¹ Hence the insight of Markowitz (1959) applies to the construction

¹ For instance, imagine that a lender knows, based on credit-risk scores, that the expected default rate on a pool of consumer receivables is 5%. This is consistent with two very different scenarios: 1) in a state of the world that occurs 5% of the time (e.g. a recession), everyone in the pool defaults, and 2) everyone in the pool has an independently-

of portfolios of consumer loans: If a lender cares about the risk that a loan *adds* to his portfolio, he should prefer loans with low covariance risk, *c.p.*

Our methodology is analogous to that of the equity-pricing literature. Using individual consumers as our unit of observation, we track their credit scores to approximate the monthly returns experienced by their lenders. We then aggregate these individual time series into a time series of aggregate returns on consumer loans, and compute each consumer's 'default beta' as the covariance of his individual returns with the aggregate returns. These default betas measure consumers' covariance risk.

After computing the default betas, we investigate two key issues. First, is there systematic heterogeneity in the betas across consumers? For instance, how does covariance risk vary with income and age and with local economic conditions? The answers to such questions help characterize the shocks hitting different types of consumers, and so are of macroeconomic interest.² The answers are also of interest from the point of view of a lender. A lender can directly compute the beta of a current borrower assuming that a suitable time series of her credit usage is available in-house. But in the absence of such a time series, a potential new borrower's beta must be estimated from her current characteristics. The second issue we investigate is whether empirically lenders put some weight on covariance risk in determining the amount of credit they

distributed 5% probability of defaulting, so in all states of the world 5% of the pool defaults. In scenario 1 the individual receivables are perfectly correlated (maximal covariance), whereas in scenario 2 they are uncorrelated. The lender would presumably prefer scenario 2, *c.p.* Since a constant 5% of the pool is defaulting in every given year, the aggregate cash flows from the pool are smooth relative to scenario 1. This makes it easier to securitize the pool, since a claim on say the first 95% of the pool's cash flows would pay 100 cents on the dollar no matter what. By contrast, a lender in scenario 1 has a problem. No claim on the pool is risk-free without complete credit-enhancement. Hence the receivables' covariance structure determines how the pool can be carved into different securities, and the type and amount of credit-enhancement required. The securitization market shows informal appreciation of this covariance risk, in that a new issue's prospectus will usually include some information regarding the diversification of the pool, such as its distribution across geographic and other characteristics (e.g., the age/seasoning of the debts).

provide. It is well known that lenders take into account a consumer's expected probability of default, in particular as measured by his credit score. The open question is whether consumers with greater covariance risk obtain less credit, even controlling for their scores and other factors. We compare the heterogeneity in covariance risk with the heterogeneity in credit risk as measured by the credit scores, and also compare the effect on credit of covariance risk with the effect of the credit scores. More generally, this is the first paper we know of to systematically study the determinants of the cross-sectional distribution of credit using the comprehensive credit bureau data.

We obtain two key sets of results. First, there is significant systematic heterogeneity across consumers in covariance risk. Covariance risk is higher for younger and single consumers, lower-income consumers, those who rent rather than own, and those from states with higher rates of divorce and lower rates of health-insurance coverage. Consumers with high covariance risk also tend to have low credit scores (high default probabilities). Second, the amount of credit obtained by consumers significantly increases with their credit scores, *c.p.* Even controlling for the credit scores and other factors, the amount of credit significantly declines with consumers' covariance risk, *c.p.* We find declines in both non-revolving debt and especially revolving debt such as bankcard credit, in the number of loans and in the dollar value of credit limits extended. It appears that some lenders do take covariance risk into account, at least in part, in determining the amount of credit they provide. Nonetheless, the effect of covariance risk on credit is much smaller in magnitude than that of the credit scores.

The rest of the paper is divided into five sections. Section I discusses how we extend asset pricing theory to consumer credit. Section II describes the dataset, and Section III our

² Souleles (2004) finds that 'aggregate' macroeconomic shocks have disparate effects across different types of

econometric methodology. Section IV reports the results, and Section V concludes.

I. Applying Asset-Pricing Theory to Consumer Credit

As Markowitz (1959) observes, the risk of a security that affects the risk borne by its investors is the risk that it adds to a well-diversified portfolio. Consequently, risk-averse investors should prefer securities whose payoffs covary relatively less with their current portfolios, *c.p.* Sharpe (1964) and Lintner (1965) show that, in equilibrium, all investors hold the value-weighted portfolio of all assets, and so they predict that investors prefer lower covariance with this portfolio. Specifically, they predict that expected returns are linear in betas.

The standard application of this covariance-pricing theory is to equities. Much of the equity market is held through well-diversified portfolios, so it is plausible that there is greater investor demand for stocks with a low covariance vis-à-vis the value-weighted portfolio of all equities (i.e., for stocks with a low CAPM beta). This theory is relatively easy to test because equity-return data are widely available. Empirical research (e.g., Fama and MacBeth (1973)) generally supports covariance-pricing: Stocks with higher CAPM betas have historically paid higher subsequent returns on average.

Our analysis applies this covariance-pricing intuition to consumer credit, where data has been more scarce. Like stocks, consumer credit is also held through well-diversified portfolios. In fact, the major components of household debt, i.e., mortgages, car loans and bankcard debt, are held almost exclusively in portfolios with thousands or even millions of loans, so by this measure there is even greater diversification in the consumer credit market than in the stock

households. Pesaran et al. (2004) analyze heterogeneity in credit-risk diversification across firms.

market. This suggests that consumer lenders might prefer lending to consumers who bring a covariance benefit to the aggregate portfolio of consumer loans.

The equity-market literature relates expected stock returns to covariances relative to the market portfolio of equities, which is only a subset of the market portfolio of all assets. Strictly speaking, this is consistent with the underlying theory under the additional simplifying assumption that equities are held by investors who hold only equities. We can motivate our application to consumer credit analogously, by assuming that consumer loans are held by agents who hold only (or largely) consumer loans. As discussed above, institutional features of the consumer lending market provide independent motivation for our application.

The equity-market literature focuses on the pricing of securities, taking the quantities of securities to be fixed. That is, the preference for lower covariance is associated with a higher price, not higher quantity. However, if quantities are endogenous we should expect adjustment of both quantities and prices in equilibrium. Both are surely endogenous even in the short run in the consumer credit market, so in this case covariance theory would predict that credit supply increases as covariance risk declines, *c.p.* Indeed, the credit bureaus collect rich data on the quantities of credit, not the prices (nor interest rates), so this is the prediction on which we focus.

To apply covariance theory to consumer credit we need to calculate the return that a consumer's lenders make on their loans to him during each month. Unlike stocks, consumer debt is not publicly marked to market. Even the credit bureaus do not directly record these returns, but we can proxy for them by using the reported time-variation in consumers' probability of repayment. The change in this probability approximates the price return on loans to the consumer.

Our dataset includes for each consumer monthly observations of their Fair Isaacs Company (FICO) credit-risk score, which is the industry-standard measure of consumers' default risk. The credit scores aim to summarize the information in each consumer's credit file regarding the probability of being seriously delinquent over the next two years, where 'seriously delinquent' means anything from 90+ days delinquent to bankrupt.³ In practice lenders rely very heavily on the scores in setting their credit policy for each consumer, often using them as summary statistics for the consumer's credit-worthiness and profitability [Moore (1996)]. Thus we let $p_{i,t}$ be the default probability implied by consumer i 's credit score as of date t .

We can use these default probabilities to mark to market a stylized consumer credit. Consider a consumer who has borrowed one dollar and promised to repay it at a future date, so that the market value of the loan on date t is approximated by $1-p_{i,t}$, the current probability of repayment. Then the return on the loan at date t is approximated by $p_{i,t-1} - p_{i,t}$, the change in the probability of repayment. Because it is more natural to analyze the change in p rather than -1 times the change in p , we analyze $p_{i,t} - p_{i,t-1}$. That is, we refer to changes in the probability of default -- and 'default betas' -- as opposed to changes in the probability of repayment. Because we will look at covariances between time series that have both been multiplied by -1 , this normalization is of no consequence.⁴

II. Data Description

This paper uses a unique, proprietary panel dataset of credit files from one of the major U.S. credit bureaus, Experian. The dataset tracks approximately 100 thousand randomly sampled

³ Gross and Souleles (2002) verify that the credit scores are very significant predictors of consumer default. Musto (2004) analyzes additional properties of the scores.

consumers monthly from 1997:03 to 2000:03, a total of 37 months. (See Musto (2004) for additional information about the dataset.)^{5,6}

Credit bureau files contain comprehensive summaries of the credit relationships ('trades') held by each consumer. For each credit trade, there are various measures of the amount of debt held and the repayment performance. The underlying data are obtained primarily from the creditors, mostly financial institutions and retail lenders. Our dataset contains the partially aggregated credit reports for each consumer that are available to and used by lenders to evaluate whether to lend to the consumer. These reports aggregate the consumer's individual trades into categories reflecting different types of credit, such as mortgages, auto loans, credit cards, etc. For each consumer-month, the dataset includes dozens of variables summarizing credit usage and delinquency, as well as credit limits where applicable. For example, for bankcards the available variables include "Total number of open bankcard trades," "Total number of bankcard trades presently 90 or more days delinquent or derogatory," and "Total (sum) of credit limit on all open bankcard trades."⁷ There are analogous variables for the other credit categories.

We merge this data with another, proprietary research dataset from Experian that contains salient individual demographic characteristics such as marital status, gender, number of children, housing status (rent vs. own), date of birth and income. This demographic information covers

⁴ More generally, we need only that the p we use be linear in default probability. As noted below, this means that the credit scores need to be transformed, because they are calibrated as a non-linear function of default probability.

⁵ The sample is a geographically stratified random sample. The dataset includes an artificially generated ID variable that allows us to cross-link the various data available for a given individual without identifying the individual.

⁶ The unit of observation in the data is an individual, not a household. Thus it is not possible to study issues like risk-sharing within households. If both members of a couple are jointly responsible for a loan, e.g. a mortgage, then the loan will appear in both of their credit files. It is not possible to directly adjust for such joint accounts in the data, though the analysis of credit allocations below will control for the sample members' demographic characteristics, including marital status and gender.

⁷ Bankcards include general purpose credit cards like Visa, Mastercard, Discover, and Optima cards, as opposed to cards from retailers. Since the unit of observation is an individual, not a trade, consumers with multiple bankcards or multiple other trades appear in the sample just once.

about 80% of the credit-file sample, though some variables are populated more than others.⁸ We further augment this individual-level data with measures of local economic conditions in the region in which each consumer lives (based on state and zip codes), such as the state unemployment and divorce rates. These individual and regional variables allow us to investigate the underlying demographic determinants of covariance risk.

Table 1 provides summary statistics for the main variables used in the analysis below. In 2000 the average consumer in our sample has a total of 6.6 currently open credit relationships (*AllTrades00*). Of these trades, 1.7 are non-revolving trades, such as mortgage, auto, and other installment loans (*Nonrevolving00*), and 4.9, the majority, are revolving trades such as credit cards (*Revolving00*). The average sum of credit limits across all open bankcards (*CardLimits00*) is about \$18 thousand. The large standard deviations around these averages imply substantial heterogeneity in all these measures of credit. Along the extensive margin, 92% of the sample have at least one open credit trade ($AllTrades00 > 0$), with 73% having non-revolving trades and 85% having revolving trades.

The resulting dataset is ideally suited for the purposes of this paper. Most notably it includes the credit-risk scores, which summarize lenders' own expectations regarding consumers' probabilities of default – the first moments p . We do not need to estimate these probabilities ourselves; we use the actual (calibrated) probabilities that the lenders themselves use. With a long time series of credit scores for each consumer, we can compute for each consumer the covariance – a second moment – of the time series of changes in his probability of default with the corresponding aggregate time series. Also, we have rich measures of the credit

⁸ This demographic information is not part of consumers' credit bureau files. It was obtained from a variety of public and propriety sources, including census and marketing databases. For example, homeownership is determined from

obtained by each consumer, so we can test whether credit supply varies with the consumer's covariance risk, even controlling for his credit score and other characteristics.

The credit score measures the likelihood of default over the next 24 months. While it is scaled from 300-850 (with larger values representing smaller likelihoods of default), it was originally calibrated from actual default rates. We invert this calibration to recover the underlying default probability units. For each consumer i and month t we produce $p_{i,t}$, the expected probability of default.⁹ Then $r_{i,t} \equiv p_{i,t} - p_{i,t-1}$ is the monthly change in the probability of default. As discussed above, with its sign reversed, $r_{i,t}$ proxies for the monthly return to lending to the consumer.

Table 1 illustrates the importance of the credit scores in the allocation of credit. The third and fourth columns split the sample at the average score in 1999:03. Consumers with the higher scores have much more credit in the subsequent year, along both the extensive and intensive margins. For instance, 99% of these consumers have at least one open credit trade, compared to 85% of the low-score consumers. The high-score consumers hold almost twice as many total trades, and over twice as many revolving trades. Their credit limits are about six times larger, \$32 thousand versus \$5.5 thousand.

The default betas that we estimate below capture the systematic component of each consumer's default risk. The remaining component is idiosyncratic. In principle a lender can diversify away this idiosyncratic risk by holding a large, diversified portfolio of loans. To gauge

tax assessor and deed information, supplemented by a model predicting homeownership as a function of consumer characteristics.

⁹ We divided the initial credit scores in 1997:03 into 20 (5 percentile) bins, and then within each bin measured the default rate over the next 24 months. As indicated by the Experian User Guide, the resulting relation is exponential. Accordingly we regressed the log of the average default rate across bins on a cubic polynomial in the average score, and use the resulting predicted probabilities of default for p . Consistent with the definition of the scores, we include in default both bankruptcy and non-bankruptcy delinquency (90+ days late), conditional on not already starting in

the potential scope for diversification of consumer loans, we first contrast the results of holding diversified and undiversified portfolios of various sizes. Figure 1 illustrates the distribution of ‘long-run returns’ across various portfolios, using the 3-year change in the probability of default, $r_{i,long} = p_{i,2000:03} - p_{i,1997:03}$. The horizontal axis records the number of consumers in each portfolio, the vertical axis measures the standard deviation of returns cross-sectionally across 100 portfolios of each size. The figure is analogous to well-known figures showing the effects of diversification across stocks, sometimes constraining stock portfolios to one country or industry [e.g., Solnik (1974)].

The (lower) line labeled ‘random’ allocates consumers randomly across the 100 portfolios. The resulting standard deviation drops quickly with portfolio size, from about 1.6% for a portfolio with only one consumer, to only .06% for a portfolio with six hundred consumers, with most of the decline coming from the first few hundred consumers. This decline suggests a large scope for diversification. Lenders do not, however, hold exactly equally weighted, random portfolios of consumer loans. For comparison, the line labeled ‘geo-undiversified’ reflects the geographically worst diversified portfolio, successively adding to each portfolio the consumer with the next closest zip code in the sample.¹⁰ The standard deviation still drops quickly with portfolio size. Nonetheless for a portfolio of size 600 it drops to only about .10%, about two-thirds larger than for the random portfolio. This difference suggests that the costs of holding

default in 1997:03. As a robustness check we also estimated a logit model of default over the next two years as a function of a cubic polynomial in the initial score. This produced very similar results to those below.

¹⁰ To construct the random portfolios, within each portfolio we randomly ranked each consumer 1-600, then formed portfolios of size 1-600 based on these ranks. We compute the mean return $r^*,long$ within each portfolio, and then compute the cross-sectional standard deviation of $r^*,long$ across the 100 portfolios. We repeat this exercise 10 times and graph the average results. To start each of the 100 geographically undiversified portfolios, we randomly sampled consumers subject to the constraint that the consumer have 600 unselected consumers around him (in zip-code space). To each of these starting portfolios we add the consumer with the closest zip code, among the remaining consumers, to the first portfolio-member’s zip code.

undiversified portfolios, even large portfolios, can be significant. More generally, these results suggest that there might be substantial systematic heterogeneity in default risk across consumers of different characteristics, including here geographic location. It remains to be seen whether lenders take this heterogeneity into account in extending credit.

III. Econometric Methodology

We begin by estimating consumers' default betas, by computing the covariance of the time series of each consumer's monthly changes in default probability with the corresponding time series for the monthly 'aggregate' changes in default probability. Specifically, for each consumer i , we run the following time-series regression:

$$r_{i,t} = \alpha_i + \beta_i r_{*t} + \varepsilon_{i,t}, \quad (1)$$

where r_{*t} is the cross-sectional average of the change in default probabilities $r_{i,t}$ within the entire sample in month t . The resulting coefficient β_i is consumer i 's 'default beta'. A larger default beta represents greater covariance risk.

Below we study how these betas affect credit supply, including the change in credit over the final year of the sample period, 1999:03-2000:03. Accordingly, to avoid endogeneity, we compute the betas in equation (eq.) (1) using data from the preceding part of the original sample period, 1997:04 through 1999:03. Reducing the beta-estimation period increases the estimation error in our betas, making it harder to find an effect of the betas on credit.¹¹ Even so, almost a fifth of the resulting betas are statistically significantly different, at the 5% level, from 1.0 (the

¹¹ To improve precision we limited the sample to consumers where the default probabilities p are available for all 24 months after the first month of the sample, i.e. 1997:04-1999:03. This retains about 94% of consumers. Hence in eq. (1) r_{it} is available for every month $t=1997:05-1999:03$ for all consumers. The results below are similar, though sometimes less significant, on including the consumers with fewer available observations of p .

average beta). This suggests that the informational content of the estimated betas is relatively high.

We also considered a number of robustness checks in computing the default betas. For instance, we computed analogous betas using the change in credit scores directly ($Score_t - Score_{t-1}$), as opposed to the change in the calibrated probabilities of default. We also used non-overlapping, two-month changes in the default probabilities ($p_t - p_{t-2}$). The conclusions below are robust to these alternatives, though sometimes somewhat weaker in magnitude using the two-month changes because of the smaller effective sample.¹²

After computing the default betas, we investigate two issues. First, is there systematic heterogeneity in the betas? Second, do lenders take covariance risk into account in determining the amount of credit they provide? We compare the results with analogous results using the credit scores.

To investigate the first issue, heterogeneity in the default betas, we estimate the following equation, cross-sectionally across consumers i :

$$\beta_i = \gamma_0 + \boldsymbol{\gamma} \mathbf{X}_i + \varepsilon_i, \tag{2}$$

where vector \mathbf{X}_i includes the key demographic characteristics of consumer i available in the dataset. These variables are flexibly specified as a series of indicator variables. To maintain sample size, instead of dropping observations with missing demographic characteristics, we include additional indicator variables for the missing values with the corresponding original indicator variables then set to zero.^{13,14}

¹² We do not include stock returns as an independent variable in eq. (1), partly due to the differences in the way that they and our changes in default probabilities are computed. The correlation between r_{*t} and monthly returns on the S&P 500 stock index is negative (recall that r_{*t} is -1 times the estimated return) but statistically insignificant.

¹³ The indicator ‘*DemographicsMis*’ in the tables refers to a subset of the auxiliary demographic dataset, constituting about 20% of total sample consumers (Table 1), for whom the individual-level demographic variables are mostly all

We sometimes augment these individual-level demographic variables with measures of local economic conditions in the region in which each consumer lives. Gross and Souleles (2002) show that such economic variables can improve predictions of consumer default, even after controlling for the credit scores. *Unemployment* measures the state-level unemployment rate. This data is available monthly. Since the betas are computed using data starting in 1997:04, we take the unemployment rate from 1997:03. This timing allows us to assess the effect of unemployment on subsequent default behavior. It is also consistent with the point of view of a lender who wishes to estimate betas for potential new customers using already available information. X_i also includes the fraction of people in the state without health insurance (*Noinsurance*) and the divorce rate in the state (*Divorce*). These variables are available only annually, and so we take them from 1997. In specifications including the regional variables, the standard errors are adjusted to allow for within-state correlation.¹⁵

Other specifications replace the regional variables with state dummy variables. These dummy variables control for all state-level geographic effects on covariance risk, including state regulations that affect the costs of default (e.g., state bankruptcy laws), the composition of jobs

unavailable. The other indicators with the suffix '*Mis*' identify additional missing demographic values for the remaining consumers for whom some demographic characteristics are usually available. (The variable for number of kids is taken from birth records and other sources that help identify the number of children when children are present, but do not directly identify households with no children. Hence '*KidsMis*' reflects both missing/uncertain values and zero children.) Since these individual demographic characteristics are still sometimes not populated, for different subsets of the sample, limiting the sample to consumers with *all* of their demographic characteristics populated would eliminate about 90% of the sample. Dropping only the smaller set of consumers with *DemographicsMis*=1 eliminates 20% of the sample, but nonetheless the results are similar to those below. (The main exception is that in Table 4 beta becomes less significant for non-revolving trades.)

¹⁴ Eq. (2) does not control for consumers' credit scores because they are likely to be simultaneously determined with the betas and hence endogenous. Eqs. (3) and (4) by contrast can use both the scores and the betas as independent variables.

¹⁵ The results below are similar using the Department of Labor's county-level unemployment rate (substituting the state-level rate when the county-level rate is unavailable). The divorce variable (which comes from the Department of Health and Human Services) is missing for a few states. To avoid dropping all observations in these states, we instead introduce a dummy variable for missing divorce data. The divorce rate is computed in per capita terms, since the only readily available normalizing variable is state population.

across occupations and industries in the state, the correlation of the state business cycle with the national business cycle, etc. The state dummies also control for all differences in average household characteristics across states. For comparison, we re-estimate eq. (2) with the credit score in 1999:03 as the dependent variable.

The second issue we examine is whether lenders take some account of covariance risk in their lending decisions, at least implicitly. To test whether default betas influence the amount of credit obtained by consumers, and to analyze the distribution of credit more generally, we estimate the following equation, again cross-sectionally across consumers i :

$$Credit_{i,00} = \gamma_0 + \gamma_1 Score_{i,99} + \gamma_2 \beta_i + \gamma_3 X_i + \varepsilon_i, \quad (3)$$

where $Credit_i$ represents various measures of the credit granted to consumer i , and $Score_i$ is i 's credit-risk score. The score summarizes the expected probability of default. Although lenders are known to rely heavily on the scores in allocating credit, since the scores are not available in traditional household datasets used in previous studies, their importance in credit supply has not been previously systematically studied. Since larger scores imply a smaller probability of default, we expect to find larger scores associated with more credit, *c.p.* We will quantify how much credit increases with the score.¹⁶

By contrast, less is known about the extent to which lenders also take covariance risk into account. They might put some weight on diversifying their loans across certain characteristics, but this weighting might be done informally, not systematically. If lenders exhibit some

¹⁶ We use in eq. (3) the scores and not the underlying probabilities of default because lenders themselves use the scores, not the probabilities directly. The conclusions below are robust to using cubic polynomials in the scores. Hence the default betas are not simply picking up nonlinearities in the effect of the scores. Eq. (3) introduces the scores linearly only for ease of exposition. Also, we report the results estimating eq. (3) by OLS, but the estimated effects are quantitatively similar using tobit models. Since the vast majority of the sample has credit relationships, the reported results emphasize the intensive margin using OLS. Some recent papers have analyzed the effects of credit scores and credit quality on mortgage debt; see e.g. Pennington-Cross and Nichols (2000), Barakova et. al. (2003), and Chomsisengphet and Elul (2004).

preference for consumers with low covariance risk, even after controlling for consumers' credit scores, the coefficient γ_2 on the default beta will be negative. On the other hand, if lenders care only about the first moment of default, then γ_2 will be insignificantly different from zero.^{17,18}

Of course, the probabilities of default, and so the credit scores, are themselves influenced by the amount of debt consumers hold. Covariance risk might also be affected by debt. To minimize endogeneity, in eq. (3) the score and the beta are lagged one year relative to the dependent variable *Credit*. Specifically, we measure *Credit* as of 2000:03 and the score as of 1999:03, while the beta was computed using information from 1997:04 to 1999:03. This timing is also consistent with the analysis below of the change in credit over the final year of the sample period, 1999:03-2000:03, which further helps avoid endogeneity.

Eq. (3) will sometimes also control for X_i , the same individual-level demographic characteristics and state dummy variables described above. Including these variables does not imply that lenders directly condition on them in setting credit supply; the variables might simply be correlated with other factors on which lenders condition, including e.g. application data that is not available in the credit bureau files.¹⁹ Nonetheless we sometimes include X_i in order to test whether any estimated effect of the default beta on credit might reflect the correlation of beta with demographic and regional characteristics. This provides a more conservative estimate of the effect of beta. If beta is significant even controlling for X_i , then beta contains additional

¹⁷ To minimize the role of outliers when the estimated betas are used as independent variables, in eqs. (3) and (4) below we drop betas with absolute values above 10. Such outliers represent only about 1-2% of the sample. In eq. (2), which has beta as the dependent variable, we do not drop the outliers in the reported results, but we have verified that the conclusions are not driven by outliers.

¹⁸ Eqs. (3) and (4) do not take into account the fact that the default betas were estimated in a previous stage, and so their reported standard error must be interpreted with some caution. Nonetheless, as noted above, measurement error in the betas could substantially attenuate their point estimate, making it on balance harder to find that they have a significant effect on credit.

information relevant for credit allocations above and beyond the heterogeneity studied in eq. (2).

On the other hand, if X_i is significant or reduces the significance of beta, it remains possible that the demographic variables in X_i are still partly picking up aspects of covariance risk that are not fully captured by the imperfectly measured beta. X_i also helps control for differences across households in the demand for credit.

The credit bureaus characterize credit relationships by type, one of the key distinctions being revolving versus non-revolving credit. Non-revolving credit consists mostly of mortgages, auto loans, and other installment loans. Such loans typically involve borrowing a fixed amount against a given purchase and repaying this amount with interest over time according to a pre-arranged schedule. By contrast, revolving loans are typically open-ended and uncollateralized. The most important example is bankcard credit cards, which give the borrower a flexible line of credit. In addition to having different contractual features, revolving and non-revolving loans can differ in the type of information available to lenders, in addition to credit bureau information, that might help them assess the covariance risk of different consumers. For both types of loans lenders can obtain additional information at the time of application, but the application information is typically richer for non-revolving loans. On the other hand, credit card issuers continue to manage their credit limits over time. In doing so they can take advantage of the rich information that they internally accumulate over time concerning the past behavior of their customers.

Accordingly, to allow for potential differences in the ability and practice of different lenders to assess and put weight on consumers' covariance risk, in eq. (3) we use different

¹⁹ Indeed, under the Equal Credit Opportunity Act lenders cannot directly condition on certain demographic characteristics, including gender, age and marital status, as well as race, religion and national origin.

measures of *Credit*₀₀. We distinguish the number of non-revolving and revolving trades. Since the latter largely reflects bankcards, we also consider the dollar-magnitude of revolving credit by studying total bankcard credit limits for each consumer. Compared to non-revolving debt, with revolving debt one can more readily distinguish credit supply from credit demand by focusing on credit limits as opposed to credit balances.²⁰

Even though the credit score and default beta used in eq. (3) are lagged, and so contain only information available by 1999:03, some of the credit relationships recorded in the dependent variable might have been initiated before 1999:03. To minimize any resulting endogeneity, we also estimate the relation between credit and beta in differences:

$$d(Credit)_{i,00} = \gamma_0 + \gamma_1 Credit_{i,99} + \gamma_2 d(Score)_{i,99} + \gamma_3 \beta_i + \gamma_4 \beta_i * d(Score)_{i,99} + \gamma_5 X_i + \varepsilon_i, \quad (4)$$

where $d(Credit)_{00}$ is the change in credit between 2000:03 and 1999:03. Working in differences also controls for all individual fixed effects in the level of credit. Further, to allow the magnitude of the change in credit to vary with the initial amount of credit, $Credit_{99}$ is the starting amount of credit in 1999:03. To allow a discontinuity for people starting without any credit, we also include an indicator ($Credit99_zero$) that equals one if the amount of credit in 1999:03 is zero.

What could prompt lenders to supply more credit to a given consumer? One of the most important factors would be an increase in her credit score. Accordingly the independent variables include the change in the score. To avoid endogeneity the change is lagged, and computed over the prior two-year period over which the beta was also computed: $d(Score)_{i,99} = Score_{99:03} - Score_{97:03}$. This timing treats the beta and the score symmetrically. It is also consistent with the point of view of a lender who, in considering in 1999 whether to extend credit to consumer i , considers variables potentially available at the time.

²⁰ E.g., see Gross and Souleles (2002a). The results for bankcard limits below are similar on also controlling for

For a given change in the credit score, the resulting change in credit might vary with the consumer's default beta. Hence eq. (4) also includes the interaction of beta with the change in score, $\beta_i * d(\text{Score})_{i,99}$, as well as the beta directly. If consumers with larger default betas receive smaller increases in credit for the same increase in their scores, then the coefficient γ_4 on the interaction term would be negative. Such a result would be especially suggestive that lenders give some consideration to covariance risk in allocating credit, since eq. (4) controls for both beta and the change in the score separately (and sometimes also for the demographic characteristics and state dummies in X_i).²¹

IV. Results

We begin by analyzing the estimated default betas. Is there systematic heterogeneity in the betas across consumers, which lenders can potentially take into account in supplying credit? We next proceed to analyze the actual supply of credit. Do consumers with greater covariance risk obtain less credit, even controlling for their expected probability of default (as measured by the credit scores) and other factors?

Table 2 presents tests for heterogeneity in the estimated default betas, following eq. (2). As a starting point, column (1) considers only individual-level demographic characteristics. The omitted demographic categories are low income (below \$25 thousand), young (below age 30), unmarried, female, one child, one adult, renter, and not a business owner. The indicator variables

bankcard balances in eq. (3), and for the change in balances in eq. (4), despite the potential endogeneity of balances.

²¹ One could additionally consider in eq. (4) the effect of a change in beta, but it cannot be readily estimated in the data. E.g., betas estimated over two-year windows will be slow to adjust over time, and 1999:03-2000:03 would be a short period for estimating the updated value of beta. Similarly in eq. (3), one could consider the effect on credit of the 'contemporaneous' beta, instrumenting for it with a lagged beta, but it is difficult to estimate a contemporaneous beta corresponding to Credit00 at the point in time 2000:03.

for missing or unknown demographic values are labeled with the suffix ‘Mis’.²² The demographic variables are jointly statistically significant, and many are individually significant. Starting at the top of the column, the default betas significantly decline with income. The beta of a high-income consumer (above \$50 thousand) is about 0.24 smaller than that of a low-income consumer, *c.p.* Relative to the average beta of 1.0, this effect is also economically significant. The betas also decline monotonically with age, with the effect large in magnitude. Consumers aged 60 and above have average betas about .42-.56 smaller than those of young consumers. Hence high-income and older consumers exhibit less covariance risk – they are less likely to default when the aggregate default rate is high, and vice versa, relative to other consumers. The betas are also smaller for consumers who are married, female, have children, or own their homes. By contrast, business ownership is insignificant, but that might be driven by the small incidence of business ownership reported in the data.

Column (2) instead considers regional economic conditions in the state in which the consumer lives. These conditions are measured as of 1997, in order to see their effect on the default betas, which were estimated over 1997-1999. The regional variables are jointly significant. The betas significantly increase with the fraction of people in the state lacking health insurance and with the divorce rate in the state. The coefficient on the state unemployment rate is also positive, though insignificant. Hence, in addition to increasing the probability of default, as previously studied, adverse regional conditions also tend to increase covariance risk. Column (3) includes both the individual and regional characteristics. The results are qualitatively similar to those in the previous columns, though the divorce rate is less significant in conjunction with the individual controls.

²² Recall that for missing values the corresponding original indicator variables are set to zero, so the coefficients on

Column (4) replaces the regional controls with state dummy variables, which control for all state-level geographic effects on covariance risk. The (unconditional) sample average default betas vary substantially across states, ranging from about .7 in Vermont and Alaska to about 1.3 in Delaware. Generally the betas tend to be larger in the South and smaller in New England.²³ In column (4) of Table 2 the state dummies (not shown) are jointly statistically significant. That is, even controlling for the available individual-level characteristics, the betas still vary significantly across states. Nevertheless the individual characteristics remain significant.

For comparison, column (5) repeats the specification of column (3) using the credit score in 1999 as the dependent variable instead of the default beta. It is important to recognize that the score does not directly condition on the individual and regional demographic control variables used here. (Indeed, the controls come from different datasets.) Instead, the results simply reflect reduced-form correlations between these demographic characteristics and the credit bureau information predictive of default that is used in constructing the scores (mostly debt holdings and repayment histories).²⁴ Because they tend to be less likely to default, high-income, older, married, and home-owning consumers have significantly larger scores. Scores also tend to be greater in states with lower rates of unemployment, divorce, and lack of health insurance, though only the latter is statistically significant.²⁵

the ‘Mis’ indicators should be interpreted relative to the omitted categories.

²³ In light of the smaller bin sizes, these unconditional results by state drop outliers in the betas, as in Tables 3 and 4.

²⁴ According to Fair Isaacs, about 35% of the score reflects measures of payment history, 30% amounts owed, 15% length of credit history, 10% new credits applied for, and 10% other factors such as the mix of credit types. The scores are not functions of demographic characteristics like gender and marital status disallowed under the Equal Credit Opportunity Act [Fair Isaacs (2005)].

²⁵ The significance of regional shocks might vary with the sample period. Our sample period ends before the 2001 recession, so there might not be enough variation in unemployment in the sample to attain significance in Table 2. By contrast, Gross and Souleles (2002) found that cross-state differences in unemployment rates were correlated with consumer bankruptcy probabilities over their 1995-97 sample period.

Notably, the effects of the demographic characteristics on the scores in column (5) usually have the opposite sign as their effects on the default betas in column (3). For instance, younger and lower-income consumers tend to have both lower scores and higher betas, and so might receive less credit for both reasons. Hence, consumers who are risky in terms of their default probability (the first moment p) also tend to have greater covariance risk (the second moment β). This could reflect the possibility that consumers who are near the edge of default (with large p) are very vulnerable to aggregate shocks, whereas those far from default might face more idiosyncratic shocks. As a result, the betas and the scores overall have a negative correlation (of about -.2 and significant), as is also evident in Table 1.²⁶

In all columns in Table 2 the R^2 is relatively small, indicating that the individual and regional characteristics explain only a small fraction of the cross-sectional variation in the credit scores and estimated default betas. This could partly reflect measurement error in the demographic characteristics. The R^2 is much smaller for the betas however, which could reflect the imprecision in their estimation.²⁷ Nonetheless the individual and regional variables are jointly statistically very significant in explaining the betas in columns (1)-(4). We conclude that there is significant systematic heterogeneity across consumers in their default betas.

Table 3 tests whether the differences in consumers' default betas affect the amount of credit they actually receive. The four pairs of columns consider the total number of credit trades, the number of non-revolving and revolving trades, and total credit card limits for each consumer, respectively. Following eq. (3), to minimize endogeneity credit is measured as of 2000

²⁶ Suppose consumer i 's probability of default can be described as $P_{it} = \delta_i + \gamma_i Y_t$, where δ_i captures his average probability of default, and γ_i his sensitivity to an aggregate state variable Y such as the business cycle (consistent with heterogeneity in the default betas). These results are consistent with δ_i and γ_i being positively correlated.

(*Credit00*), whereas the credit score is taken from 1999 and the default beta is computed over 1997-1999. Table 3 begins with the total number of trades, *AllTrades00*. Column (1) includes as independent variables only the credit score and beta. The score has a large, positive effect on the number of trades. This effect is highly significant, with a t-ratio above 100. Cross-sectionally, increasing the score by one standard deviation (about 0.9) increases the number of trades on average by about 2 ($\approx .9*2.095$), *c.p.* Since the average number of trades is under 7, this effect is also economically quite significant. As expected, credit scores play an important role in the allocation of credit. The coefficient on the default beta is also significant, with a negative sign. Consumers with greater covariance risk do in fact obtain less credit, even controlling for their expected probability of default. Cross-sectionally, a one standard deviation increase in beta (about 2.5 in this sample) reduces the total number of trades by about 0.22. While this effect is much smaller than that due to the score, it is nonetheless statistically significant.

Column (2) adds the individual-level demographic characteristics and state dummies (not shown). The individual and state variables are each jointly significant, though less so than the credit score. *Ceteris paribus*, the number of trades increases with income, and exhibits an inverted-U shape life-cycle profile in age, peaking in the 50s. These effects are large in magnitude, with high-income and middle-aged (50-59) consumers having about 1.3 more trades on average than low-income and young consumers. The number of trades is also larger for consumers who are married and have larger households. Homeowners and business-owners also hold significantly more trades.

²⁷ Tables 3 and 4 below will control for the same demographic characteristics as in Table 2. Hence the significance of beta in Tables 3 and 4 implies that the variation in beta unexplained in Table 2 cannot reflect only measurement error.

Adding the demographic characteristics and state dummies increases the R^2 somewhat relative to column (1), but as already explained this does not imply that lenders directly condition on these variables. Moreover these variables can of course also capture differences across consumers in the demand for credit, not just in the supply of credit. Nonetheless, controlling for the individual and state variables has little effect on the coefficients for beta and the score.

Columns (3) and (4) show analogous results for only non-revolving trades. The pattern of coefficients is qualitatively similar to that for all trades, except for gender and an earlier peak in the effect of age. Beta again has a significant negative effect.

Columns (5) and (6) instead isolate only revolving trades. The results are quantitatively similar to those in the first two columns for all trades. This is not surprising since the majority of total trades are revolving. The coefficient on beta for revolving trades (column (5)) is almost four times that for non-revolving trades (column (3)), a somewhat larger effect than the ratio of revolving to non-revolving trades. Moreover, the R^2 for revolving trades is much larger. While the demographic controls provide relatively less incremental explanatory power for revolving trades (columns (6) relative to (5) versus columns (4) relative to (3)), the default betas and especially the credit scores are relatively more significant in explaining revolving trades (columns (5) versus (3)).

Columns (7) and (8) instead consider total credit card limits. In column (7) the credit score has a significant and large, positive effect on credit. Cross-sectionally, increasing the score by one standard deviation increases average credit limits by about \$13 thousand. The default beta again has a significant negative effect. A one standard deviation increase in beta reduces credit limits by almost \$900, a substantial amount. Column (8) adds the individual characteristics and state dummies, which again are each jointly significant. The pattern of signs on the coefficients is

qualitatively similar to that in column (2). The demographic effects are often large in magnitude. The credit limits of high-income and middle-aged consumers are \$8-12 thousand larger on average than those of low-income and young consumers, *c.p.* Homeowners and business owners also have relatively large credit limits.

Overall, the results of Table 3 suggest that lenders put some weight, at least implicitly, on covariance risk in their decisions about credit allocation, though much less weight than on the credit scores. Moreover, given the negative correlation between the score and the default beta (and the imprecision in estimating the beta), by regulating the score lenders also indirectly regulate their covariance risk at the same time. To illustrate the potential effect of the correlation between the two measures of risk, one can re-estimate eq. (3) without the credit score. This significantly increases the magnitude of the estimated coefficients on beta, by almost half for non-revolving debt in column (3) and by about three times for revolving debt in columns (5) and (7).

Table 4 instead investigates the relation between credit and beta in differences, following eq. (4). In columns (1) and (2) the dependent variable is the change between 1999-2000 in the number of total trades, $d(All\ Trades)00$. To control for factors motivating the change in credit, we control for the change in the consumer's credit score over the previous period ($d(Creditscore)99$), and focus on the interaction of beta with the change in score ($d(Score)99*Beta99$). For a given increase in the score, we test whether the resulting increase in credit is smaller for consumers with a larger default beta.

The dependent variables include the lagged, starting number of trades in 1999 ($Credit99$), as well as an indicator ($Credit99_zero$) that equals one if the number of trades in 1999 starts at zero. In column (1), the lagged number of trades has a significant negative coefficient. Thus the

increase in credit is smaller for consumers starting with a greater amount of credit, suggesting some mean reversion in the reported amount of credit across consumers. On the other hand, the indicator for zero credit also has a significant negative coefficient, consistent with a potential discontinuity at zero: Consumers starting with no credit receive relatively less credit going forward.

The coefficient on the change in the credit score is significantly positive, with a large t -ratio of about 40. To interpret its effect, consider a consumer with a beta of zero. If her score increases by .6 (the cross-sectional standard deviation in $d(\text{Creditscore})$), over the next year this would boost her number of trades by, on average, $(0.6)(0.628) \approx .38$, *c.p.* Even when estimated in differences using only within-consumer variation, the score continues to have a substantial effect on the amount of credit, even within a short period of 12 months. Nonetheless the coefficient on the default beta is significantly negative, implying that consumers with greater covariance risk on average experience smaller increases in credit over time. Notably, the interaction of beta with the change in score has a significant negative effect. For a given increase in the score, consumers with higher covariance risk do in fact receive a smaller increase in credit. Consider for comparison a second consumer, same as the first except that his beta is not zero but 2.5, one standard deviation higher. Even with no change in his credit score, the growth in his number of trades would be lower on average by $(2.5)(0.019) \approx 0.048$, *c.p.* If his score increases by .6, this would have the additional effect of boosting his number of trades by, on average, $.38 - (0.6)(2.5)(0.029) \approx 0.333$, *c.p.* So overall after a .6 increase in score, the second consumer would experience over the next year a net increase in trades of $0.333 - 0.048 \approx 0.29$, which is about 25% less than the first consumer facing the same increase in score.

Column (2) adds the individual demographic characteristics and state dummies, which are again each jointly significant. For instance, the change in the number of trades between 1999 and 2000 is on average greater for higher-income consumers, but smaller for older consumers. Nonetheless the coefficients on both beta and the interaction of beta with the change in score remain significant and negative. Moreover, the R^2 starts large at .31 in column (1) and increases very little in column (2). Hence, after differencing out fixed effects in the level of credit, much more of the variation in the change in credit is accounted for by the covariates in column (1), especially the change in score and its interaction with beta, than by the demographic controls.

In columns (3) and (4), for non-revolving trades, while the effects of beta and its interaction term remain negative in sign, the interaction term is less statistically significant, especially in column (4). By contrast, in columns (5) and (6) for revolving trades, both beta and its interaction term are significantly negative. The patterns of the other coefficients in columns (3)-(6) are generally qualitatively similar to those in columns (1) and (2).

In column (7) for total bankcard limits, beta and the interaction term again have significant negative effects. For a consumer with zero beta, if her score increases by .6, over the next year her credit limits would increase on average by about \$1000, *c.p.* For an otherwise similar consumer with a one standard deviation larger beta, the resulting increase in limits, including both the interacted and uninteracted effect of beta, would be only about \$200, substantially smaller. In column (8) the pattern of demographic coefficients is generally qualitatively similar to that in column (2), with the exception that the effect of age is positive.

Overall, the results of Table 4 imply that consumers with greater covariance risk obtain significantly less credit over time. Covariance risk appears to be particularly important for

revolving credit. This could reflect the fact that providers of revolving credit have available a rich prior performance history for their borrowers.

V. Conclusion

This paper takes a portfolio view of consumer credit. It uses a unique credit bureau dataset to estimate the covariance risk of individual consumers and to analyze its underlying demographic determinants, as well as more generally the determinants of the cross-sectional distribution of credit, including the credit scores.

We obtain two key sets of results. First, there is significant systematic heterogeneity across consumers in covariance risk. Covariance risk tends to be higher for younger and single consumers, lower-income consumers, those who rent rather than own, and those from states with higher rates of divorce and lower rates of health-insurance coverage. Consumers with high covariance risk also tend to have low credit scores (high default probabilities). Second, the amount of credit obtained by consumers significantly increases with their credit scores, and, especially for revolving credit, significantly decreases with their covariance risk. This conclusion holds both in levels (in the cross-section) and in differences (for given consumers over time). It appears that some lenders assess borrowers' covariance risk, at least implicitly, and put some weight on it in determining the amount of credit they provide. Nonetheless the effect of covariance risk on credit is much smaller in magnitude than that of the credit scores, suggesting the possibility that lenders might benefit from more systematic and quantitative consideration of covariance risk.

References

- Barakova, I., R. W. Bostic, P.S. Calem, and S.M. Wachter, 2003, "Does Credit Quality Matter for Homeownership," *Journal of Housing Economics*, 12, pp. 318-336.
- Chomsisengphet, S. and R. Elul, 2004, "Bankruptcy Exemptions, Credit History, and the Mortgage Market," working paper, Federal Reserve Bank of Philadelphia, 04-14R.
- Fair Isaacs, 2005, "Your Credit Scores."
- Fama, E. F., and J. D. MacBeth, 1973, "Risk, Return and Equilibrium: Empirical Tests," *Journal of Political Economy*, 71, May-June, pp. 607-636.
- Federal Reserve System, Board of Governors, 2001, *Flow of Funds Accounts of the United States*.
- Gorton, G., and N.S. Souleles, 2004, "Securitization and Special Purpose Vehicles," working paper, University of Pennsylvania.
- Gross, D., and N.S. Souleles, 2002, "An Empirical Analysis of Personal Bankruptcy and Delinquency," *Review of Financial Studies*, 15(1), Spring, pp. 319-347.
- Gross, D., and N.S. Souleles, 2002a, "Do Liquidity Constraints and Interest Rates Matter for Consumer Behavior? Evidence from Credit Card Data," *Quarterly Journal of Economics*, 117(1), February, pp. 149-185.
- Lintner, J., 1965, "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *Review of Economics and Statistics*, 47, February, pp. 13-37.
- Markowitz, H.M., 1959, *Portfolio Selection: Efficient Diversification of Investments*. John Wiley, New York.
- Moore, M., 1996, "Credit Scoring's Uses Expand as It Gains Acceptance," *The American Banker*, p. 4A.
- Musto, D.K., 2004, "What Happens when Information Leaves a Market? Evidence from Post-Bankruptcy Consumers," *Journal of Business*, 77(4), pp. 725-748.
- Pennington-Cross, A., and J. Nichols, "Credit History and the FHA-Conventional Choice," *Real Estate Economics*, 28(2), pp. 307-336.
- Pesaran, M. H., T. Schuermann, and B.-J. Treutler, 2004, "The Role of Industry, Geography and Firm Heterogeneity in Credit Risk Diversification," working paper.

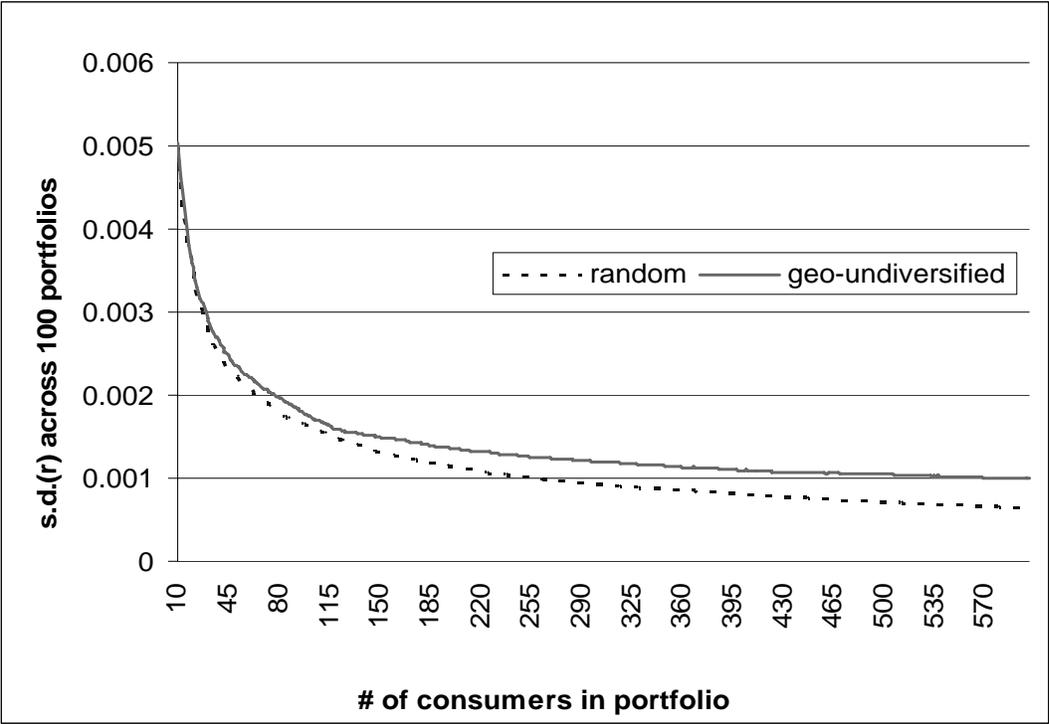
Saunders, A., 1999, *Credit Risk Management*. John Wiley & Sons, New York.

Sharpe, W. F., 1964, "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, 19, September, pp. 425-442.

Solnik, B.H., 1974, "Why Not Diversify Internationally Rather Than Domestically?" *Financial Analysts Journal*, July-August, pp. 48-54.

Souleles, N.S., 2004, "Expectations, Heterogeneous Forecast Errors, and Consumption: Micro Evidence from the Michigan Consumer Sentiment Surveys," *Journal of Money, Credit, and Banking*, 36(1), February, pp. 39-72.

Figure 1: Diversified vs. Undiversified Credit Portfolios



Notes: The vertical axis records the standard deviation of 'long-run returns' $r_{,long}$, cross-sectionally across 100 portfolios of the size indicated on the horizontal axis (averaged over 10 replications). The random portfolio adds consumers randomly in building the portfolios, the geo-undiversified portfolio successively adds the consumers in the sample with the next closest zip code. For clarity, the horizontal axis shows portfolios of size 10 and above.

Table 1: Summary Statistics

	Mean	S.D.	Creditscore99	
			Low	High
AllTrades00	6.66	5.32	4.59	8.79
Nonrevolving00	1.77	1.80	1.55	1.99
Revolving00	4.89	4.53	3.05	6.80
CardLimits00 (\$1000)	18.65	32.97	5.53	32.19
AllTrades00 >0	0.92	0.28	0.85	0.99
Nonrevolving00 >0	0.73	0.44	0.66	0.80
Revolving00 >0	0.85	0.36	0.73	0.98
CardLimits00 >0	0.75	0.43	0.59	0.92
Creditscore99	6.45	0.92	5.66	7.25
Beta99	1.00	2.96	1.50	0.49
Income25-50	0.55	0.50	0.59	0.51
Income50+	0.30	0.46	0.23	0.36
IncomeMis	0.01	0.09	0.01	0.01
Age30_39	0.19	0.39	0.20	0.18
Age40_49	0.19	0.39	0.18	0.20
Age50_59	0.13	0.34	0.11	0.16
Age60_69	0.06	0.24	0.04	0.08
Age70+	0.04	0.21	0.03	0.06
AgeMis	0.10	0.30	0.10	0.09
Married	0.31	0.46	0.25	0.37
MarriedMis	0.13	0.34	0.13	0.14
SexMale	0.50	0.50	0.49	0.51
SexMis	0.25	0.43	0.22	0.27
Kids2	0.05	0.22	0.04	0.06
Kids3+	0.03	0.16	0.02	0.03
KidsMis	0.61	0.49	0.59	0.64
Adults2	0.24	0.43	0.22	0.26
Adults3+	0.35	0.48	0.31	0.40
Homeown	0.40	0.49	0.31	0.49
HomeownMaybe	0.16	0.37	0.16	0.17
HomeownMis	0.21	0.41	0.26	0.16
Busown	0.011	0.11	0.009	0.014
DemographicsMis	0.20	0.40	0.25	0.16
Unemployment	5.10	1.01	5.14	5.05
Noinsure	16.74	4.49	17.13	16.34
Divorce	0.40	0.19	0.41	0.39
DivorceMis	0.12	0.33	0.12	0.13

Notes: The sample corresponds to that in Table 2. (N = 87,013, but sample size for individual variables may vary with missing values.) The omitted demographic categories are low income (below \$25 thousand), young (below 30), unmarried, female, one child, one adult, renter, and not a business owner. 'DemographicsMis' is a dummy variable indicating that most demographic variables are missing. Additional missing or unknown demographic values are labeled with the suffix 'Mis'. 'KidsMis' reflects both missing values and 0 children. HomeownMaybe is predicted homeownership. State divorce rate is in per

capita terms. Creditscore is the FICO credit bureau score/100. Credit limits are in \$1000s. *AllTrades00 >0* is an indicator for having a positive numbers of trades, etc. Columns (3) and (4) split the sample at the average credit score of 6.45.

Table 2: Heterogeneity in the Default Beta

	(1) Individual Demographics			(2) Regional Demographics			(3) Individual and Regional Demographics			(4) Individual Demographics w/ State Dummies		
	coef	s.e.	p-val	coef	s.e.	p-val	coef	s.e.	p-val	coef	s.e.	p-val
Income25-50	-0.084	0.030	0.005				-0.086	0.033	0.011	-0.078	0.030	0.010
Income50+	-0.235	0.033	0.000				-0.227	0.039	0.000	-0.247	0.034	0.000
IncomeMis	-0.318	0.113	0.005				-0.131	0.126	0.304	-0.201	0.116	0.082
Age30_39	-0.123	0.043	0.004				-0.130	0.044	0.005	-0.125	0.043	0.003
Age40_49	-0.165	0.043	0.000				-0.167	0.040	0.000	-0.168	0.043	0.000
Age50_59	-0.189	0.046	0.000				-0.192	0.044	0.000	-0.190	0.046	0.000
Age60_69	-0.422	0.054	0.000				-0.437	0.046	0.000	-0.429	0.054	0.000
Age70+	-0.558	0.060	0.000				-0.562	0.060	0.000	-0.569	0.060	0.000
AgeMis	-0.185	0.048	0.000				-0.190	0.052	0.001	-0.197	0.048	0.000
Married	-0.137	0.033	0.000				-0.129	0.030	0.000	-0.124	0.033	0.000
MarriedMis	-0.017	0.040	0.681				-0.018	0.032	0.584	-0.018	0.040	0.650
SexMale	0.062	0.027	0.021				0.065	0.034	0.058	0.062	0.027	0.019
SexMis	0.020	0.034	0.552				0.025	0.043	0.554	0.020	0.034	0.551
Kids2	-0.114	0.055	0.037				-0.105	0.039	0.010	-0.107	0.055	0.051
Kids3+	-0.070	0.069	0.314				-0.057	0.081	0.484	-0.064	0.069	0.355
KidsMis	-0.021	0.035	0.534				-0.025	0.033	0.464	-0.020	0.035	0.570
Adults2	0.013	0.037	0.726				0.001	0.037	0.978	0.004	0.037	0.909
Adults3+	0.033	0.039	0.390				0.022	0.032	0.504	0.019	0.039	0.623
Homeown	-0.203	0.070	0.004				-0.210	0.102	0.045	-0.190	0.071	0.007
HomeownMaybe	-0.069	0.072	0.343				-0.074	0.111	0.505	-0.050	0.073	0.490
HomeownMis	0.063	0.072	0.381				0.049	0.114	0.669	0.076	0.072	0.294
Busown	-0.004	0.095	0.970				-0.020	0.084	0.817	-0.019	0.095	0.839
DemographicsMis	-0.088	0.087	0.315				-0.100	0.112	0.374	-0.089	0.088	0.309
Unemployment				0.022	0.019	0.257	0.024	0.019	0.222			
Noinsure				0.015	0.004	0.001	0.013	0.004	0.002			
Divorce				0.358	0.109	0.002	0.255	0.137	0.069			
DivorceMis				0.239	0.069	0.001	0.197	0.078	0.015			
cons	1.379	0.088	0.000	0.468	0.123	0.000	0.935	0.170	0.000			
(Adj.) R-squared		0.0059			0.0011			0.0069			0.0082	
Obs.		87,013			85,961			85,961			87,013	

Table 2: Heterogeneity in the Default Beta (ctd)

(5)			
Credit Score on Demographics			
	coef	s.e.	p-val
Income25-50	0.127	0.016	0.000
Income50+	0.340	0.021	0.000
IncomeMis	0.293	0.048	0.000
Age30_39	0.121	0.012	0.000
Age40_49	0.194	0.014	0.000
Age50_59	0.357	0.021	0.000
Age60_69	0.573	0.020	0.000
Age70+	0.726	0.030	0.000
AgeMis	0.233	0.017	0.000
Married	0.110	0.008	0.000
MarriedMis	0.007	0.008	0.440
SexMale	-0.016	0.013	0.224
SexMis	0.015	0.007	0.041
Kids2	0.060	0.019	0.002
Kids3+	0.061	0.027	0.026
KidsMis	0.028	0.013	0.034
Adults2	0.012	0.012	0.344
Adults3+	-0.036	0.008	0.000
Homeown	0.230	0.021	0.000
HomeownMaybe	0.125	0.022	0.000
HomeownMis	-0.085	0.024	0.001
Busown	-0.030	0.027	0.268
DemographicsMis	0.154	0.031	0.000
Unemployment	-0.004	0.012	0.701
Noinsure	-0.015	0.002	0.000
Divorce	-0.067	0.112	0.550
DivorceMis	-0.083	0.066	0.218
cons	6.223	0.071	0.000
(Adj.) R-squared		0.1058	
Obs.		85,961	

Notes: See eq. (2). In columns (1)-(4), the dependent variable is the Default Beta, computed over 1997:04-1999:03. In column (5), it is the Creditscore, from 1999:03. See Table 1 for description of the independent variables. Regional demographics are taken from 1997 (3/1997 for unemployment). Column (4) includes state dummies (not shown). Columns (2), (3) and (5), with state demographics, adjust the standard errors for within-state correlation and report the unadjusted R-squared.

Table 3: The Amount of Credit and the Default Beta -- Levels

Credit00	All Trades00						Nonrevolving Trades00					
	(1)			(2)			(3)			(4)		
				w/ demographic controls						w/ demographic controls		
	coef	s.e.	p-val	coef	s.e.	p-val	coef	s.e.	p-val	coef	s.e.	p-val
Creditscore99	2.095	0.019	0.000	1.859	0.019	0.000	0.152	0.007	0.000	0.110	0.007	0.000
Beta99	-0.087	0.007	0.000	-0.083	0.007	0.000	-0.019	0.002	0.000	-0.017	0.002	0.000
Income25-50				0.550	0.050	0.000				0.136	0.018	0.000
Income50+				1.354	0.056	0.000				0.353	0.020	0.000
IncomeMis				0.867	0.189	0.000				0.077	0.068	0.259
Age30_39				0.466	0.070	0.000				0.072	0.025	0.005
Age40_49				1.034	0.070	0.000				0.074	0.025	0.004
Age50_59				1.338	0.075	0.000				-0.007	0.027	0.789
Age60_69				0.398	0.090	0.000				-0.436	0.033	0.000
Age70+				-1.000	0.099	0.000				-0.868	0.036	0.000
AgeMis				0.708	0.079	0.000				0.158	0.029	0.000
Married				0.247	0.054	0.000				0.093	0.020	0.000
MarriedMis				-0.078	0.066	0.235				-0.065	0.024	0.006
SexMale				-0.415	0.044	0.000				0.245	0.016	0.000
SexMis				0.005	0.056	0.924				-0.050	0.020	0.014
Kids2				0.350	0.089	0.000				0.065	0.032	0.045
Kids3+				0.265	0.113	0.019				-0.028	0.041	0.485
KidsMis				-0.198	0.057	0.000				-0.073	0.020	0.000
Adults2				0.115	0.061	0.061				0.067	0.022	0.003
Adults3+				0.203	0.064	0.001				0.070	0.023	0.002
Homeown				1.424	0.116	0.000				0.911	0.042	0.000
HomeownMaybe				0.290	0.119	0.015				0.386	0.043	0.000
HomeownMis				0.015	0.118	0.897				0.314	0.043	0.000
Busown				1.106	0.156	0.000				0.307	0.057	0.000
DemographicsMis				0.340	0.144	0.018				0.314	0.052	0.000
cons	-7.073	0.122	0.000				0.808	0.045	0.000			
Adj. R-squared		0.136			0.189			0.007			0.073	
Obs.		85,903			85,903			85,903			85,903	

Table 3: The Amount of Credit and the Default Beta -- Levels (ctd)

Credit00	Revolving Trades00						Credit Card Limits00					
	(5)			(6)			(7)			(8)		
				w/ demographic controls						w/ demographic controls		
	coef	s.e.	p-val	coef	s.e.	p-val	coef	s.e.	p-val	coef	s.e.	p-val
Creditscore99	1.943	0.016	0.000	1.749	0.016	0.000	14.277	0.116	0.000	12.509	0.119	0.000
Beta99	-0.069	0.006	0.000	-0.066	0.006	0.000	-0.349	0.042	0.000	-0.329	0.041	0.000
Income25-50				0.415	0.042	0.000				2.013	0.307	0.000
Income50+				1.002	0.047	0.000				8.174	0.346	0.000
IncomeMis				0.789	0.160	0.000				1.514	1.166	0.194
Age30_39				0.394	0.059	0.000				2.428	0.432	0.000
Age40_49				0.960	0.059	0.000				7.530	0.434	0.000
Age50_59				1.345	0.064	0.000				12.615	0.466	0.000
Age60_69				0.834	0.076	0.000				9.266	0.554	0.000
Age70+				-0.132	0.083	0.113				2.833	0.608	0.000
AgeMis				0.551	0.067	0.000				6.538	0.490	0.000
Married				0.153	0.046	0.001				1.506	0.334	0.000
MarriedMis				-0.013	0.056	0.816				-0.247	0.408	0.545
SexMale				-0.660	0.037	0.000				2.923	0.270	0.000
SexMis				0.055	0.047	0.241				0.901	0.343	0.009
Kids2				0.285	0.075	0.000				0.824	0.550	0.134
Kids3+				0.293	0.095	0.002				1.611	0.695	0.020
KidsMis				-0.126	0.048	0.009				-0.887	0.349	0.011
Adults2				0.048	0.052	0.354				-0.152	0.377	0.686
Adults3+				0.133	0.054	0.013				0.878	0.393	0.025
Homeown				0.513	0.098	0.000				4.731	0.715	0.000
HomeownMaybe				-0.095	0.101	0.345				1.900	0.736	0.010
HomeownMis				-0.299	0.100	0.003				0.191	0.729	0.794
Busown				0.798	0.132	0.000				7.269	0.963	0.000
DemographicsMis				0.026	0.121	0.832				5.423	0.887	0.000
cons	-7.567	0.104	0.000				-73.06	0.762	0.000			
Adj. R-squared		0.159		0.203			0.159			0.201		
Obs.		85,903		85,903			85,903			85,903		

Notes: See eq. (3). The dependent variable, Credit00, is from 2000:03. The Creditscore is from 1999:03, and Beta is computed over 1997:04 - 1999:03. Demographic controls include state dummies.

Table 4: The Amount of Credit and the Default Beta -- Changes

dCredit00	d(All Trades)00						d(Nonrevolving Trades)00					
	(1)			(2)			(3)			(4)		
				w/ demographic controls						w/ demographic controls		
	coef	s.e.	p-val	coef	s.e.	p-val	coef	s.e.	p-val	coef	s.e.	p-val
Credit99	-0.253	0.001	0.000	-0.258	0.001	0.000	-0.280	0.003	0.000	-0.292	0.003	0.000
Credit99_zero	-0.285	0.039	0.000	-0.224	0.038	0.000	-0.081	0.012	0.000	-0.043	0.012	0.000
d(Creditscore)99	0.628	0.016	0.000	0.627	0.016	0.000	0.119	0.008	0.000	0.113	0.008	0.000
Beta99	-0.019	0.003	0.000	-0.017	0.003	0.000	-0.004	0.002	0.008	-0.004	0.002	0.021
d(Score)99*Beta99	-0.029	0.005	0.000	-0.027	0.005	0.000	-0.005	0.002	0.057	-0.004	0.002	0.105
Income25-50				0.152	0.026	0.000				0.051	0.012	0.000
Income50+				0.181	0.029	0.000				0.122	0.014	0.000
IncomeMis				0.089	0.097	0.361				-0.034	0.047	0.473
Age30_39				-0.093	0.036	0.010				-0.058	0.017	0.001
Age40_49				-0.039	0.036	0.282				-0.065	0.017	0.000
Age50_59				-0.030	0.039	0.435				-0.109	0.019	0.000
Age60_69				-0.215	0.046	0.000				-0.259	0.022	0.000
Age70+				-0.583	0.050	0.000				-0.383	0.024	0.000
AgeMis				0.019	0.041	0.646				0.001	0.020	0.941
Married				0.012	0.028	0.670				0.022	0.013	0.105
MarriedMis				-0.093	0.034	0.006				-0.036	0.016	0.030
SexMale				0.220	0.022	0.000				0.075	0.011	0.000
SexMis				0.123	0.029	0.000				0.009	0.014	0.511
Kids2				0.061	0.046	0.182				-0.004	0.022	0.854
Kids3+				0.001	0.058	0.986				0.003	0.028	0.928
KidsMis				-0.006	0.029	0.842				-0.035	0.014	0.012
Adults2				0.093	0.031	0.003				0.027	0.015	0.079
Adults3+				0.106	0.033	0.001				0.022	0.016	0.168
Homeown				0.175	0.060	0.003				0.248	0.029	0.000
HomeownMaybe				0.086	0.061	0.162				0.122	0.030	0.000
HomeownMis				0.068	0.061	0.265				0.093	0.029	0.002
Busown				0.104	0.080	0.193				0.112	0.039	0.004
DemographicsMis				-0.190	0.074	0.010				-0.008	0.036	0.821
cons	0.693	0.016	0.000				0.426	0.008	0.000			
Adj. R-squared		0.305			0.315			0.153			0.165	
Obs.		85,704			85,704			85,704			85,704	

Table 4: The Amount of Credit and the Default Beta -- Changes (ctd)

dCredit00	d(Revolving Trades)00						d(Credit Card Limits)00					
	(5)			(6)			(7)			(8)		
				w/ demographic controls						w/ demographic controls		
	coef	s.e.	p-val	coef	s.e.	p-val	coef	s.e.	p-val	coef	s.e.	p-val
Credit99	-0.264	0.001	0.000	-0.266	0.001	0.000	-0.092	0.002	0.000	-0.102	0.002	0.000
Credit99_zero	-0.062	0.024	0.010	-0.015	0.024	0.527	-3.506	0.140	0.000	-2.917	0.142	0.000
d(Creditscore)99	0.491	0.013	0.000	0.494	0.013	0.000	1.719	0.106	0.000	1.714	0.105	0.000
Beta99	-0.019	0.003	0.000	-0.017	0.003	0.000	-0.228	0.023	0.000	-0.213	0.023	0.000
d(Score)99*Beta99	-0.026	0.004	0.000	-0.025	0.004	0.000	-0.170	0.033	0.000	-0.152	0.033	0.000
Income25-50				0.114	0.021	0.000				0.759	0.172	0.000
Income50+				0.094	0.024	0.000				1.945	0.194	0.000
IncomeMis				0.152	0.080	0.057				0.754	0.653	0.248
Age30_39				-0.018	0.030	0.544				0.455	0.242	0.060
Age40_49				0.053	0.030	0.076				1.190	0.243	0.000
Age50_59				0.114	0.032	0.000				1.863	0.262	0.000
Age60_69				0.069	0.038	0.068				1.891	0.309	0.000
Age70+				-0.190	0.041	0.000				0.496	0.338	0.142
AgeMis				0.041	0.033	0.217				1.364	0.274	0.000
Married				-0.001	0.023	0.973				0.312	0.187	0.096
MarriedMis				-0.057	0.028	0.041				-0.369	0.228	0.106
SexMale				0.141	0.018	0.000				0.626	0.151	0.000
SexMis				0.109	0.023	0.000				0.286	0.192	0.136
Kids2				0.072	0.038	0.054				0.742	0.307	0.016
Kids3+				0.002	0.047	0.959				0.844	0.389	0.030
KidsMis				0.026	0.024	0.271				-0.240	0.195	0.218
Adults2				0.069	0.026	0.008				0.186	0.211	0.378
Adults3+				0.086	0.027	0.001				0.309	0.220	0.160
Homeown				-0.026	0.049	0.599				1.163	0.401	0.004
HomeownMaybe				-0.018	0.050	0.727				0.721	0.412	0.080
HomeownMis				-0.021	0.050	0.674				-0.391	0.409	0.339
Busown				0.010	0.066	0.881				-0.396	0.539	0.463
DemographicsMis				-0.159	0.061	0.009				0.344	0.497	0.489
cons	0.400	0.013	0.000				4.461	0.081	0.000			
Adj. R-squared		0.349			0.358			0.031			0.039	
Obs.		85,704			85,704			85,704			85,704	

Notes: See eq. (4). The dependent variable, $dCredit00$, measures the change in credit over 1999:03-2000:03. The initial amount of credit $Credit99$ is from 1999:03, and $Credit99_zero$ is a dummy variable indicating no credit in 1999:03. The $Creditscore$ is from 1999:03, and the $Default\ Beta$ is computed over 1997:04 - 1999:03. Demographic controls include state dummies.