Online Appendix A: Inference from the relation between interest rate and ex-post defaults

This appendix formally derives what determines the relationship between interest rates and *ex-post* default. In particular, it shows that the slope of this relationship does not depend on the precision with which lenders can predict future defaults. The appendix further shows that the slope of *ex-post* defaults with respect to the interest rate becomes steeper if lenders systematically underreact to available information and that it becomes flatter if lenders base interest rates on factors unrelated to default probabilities.

To establish the determinants of the slope of *ex-post* defaults with respect to interest rates, we develop a stylized model that takes lenders to be approximately risk neutral with respect to idiosyncratic default risks over small bets (typically around $100) in the Prosper online marketplace.[1] Lenders thus require an interest rate r such that their expected returns are equal to the "risk-free" rate of return, r_f:

$$(1 + r_f) = \mathrm{E}[(1+r)(1-(1-\kappa)W)], \tag{A.1}$$

where W is an indicator variable that equals 1 if the loan defaults and $\kappa \in [0, 1)$ denotes the fraction of principal and interest that the lender still receives in case of default. Solving equation (A.1) for the expected default rate yields the expected default rate as a function of the interest rate:

$$\mathrm{E}[W \mid r] = \frac{1}{1-\kappa} - \frac{1+r_f}{1-\kappa}(1+r)^{-1}. \tag{A.2}$$

This expression indicates that we should expect a linear relationship between observed default probabilities and $(1+r)^{-1}$. In particular, if lenders are rational, the regression coefficient on $(1+r)^{-1}$ in a regression with default realizations as the dependent variable can be interpreted as $-(1 + r_f)/(1-\kappa)$. Note that the relationship in (A.2) only requires that lenders are rational and maximize expected returns, but does not depend on the precision with which lenders form the expectation of the default probability. Intuitively, if lenders have very little information on which to base their expectations, the variance in $(1+r)^{-1}$ will be low. In that case, however, the variance in $\mathrm{E}[W|r]$ will be low as well, so that the slope of the relationship between these two variables is not affected. This implies that the slope of the relationship between $\mathrm{E}[W|r]$ and $(1+r)^{-1}$ cannot inform us about the strength of lenders' signals of the default probability.

To show more formally that the relationship between $\mathrm{E}[W|r]$ and $(1+r)^{-1}$ does not depend on the precision of rational lenders' perceptions of loan default probabilities, we now model the process by which lenders assess loan default probabilities. Suppose lenders get a signal s of the true default probability of a loan. Assume that the signal equals the true (unobservable) default probability with probability π and that the signal is completely uninformative (i.e., a random draw from the distribution of true default probabilities) with probability $1-\pi$, so:

$s = \omega$ with probability π, and
$s \sim f(\omega)$ with probability $1-\pi$,

[1] The risk premium for *systematic* default risk is included in r_f.

where ω denotes the true, but unobservable, default probability of a loan and $f(\omega)$ denotes the population distribution of true default probabilities. Given this information structure, the true default probability conditional on observing signal s is:

$$E[W|s] = \pi s + (1-\pi)\ \omega_m, \tag{A.3}$$

where ω_m is the unconditional mean default probability. To derive the interest rate as a function of the observed signal, we take the expectation in the no-arbitrage condition with respect to signal s:

$$(1 + r_f) = E[(1+r)(1-(1-\kappa)W)\,|\,s]. \tag{A.4}$$

Solving (A.4) with respect to the interest rate yields the interest rate as a function of the signal:

$$r(s) = [1/(1 + r_f) - (\pi s + (1-\pi)\ \omega_m)(1-\kappa)/(1 + r_f)]^{-1} - 1. \tag{A.5}$$

We invert equation (A.5) to find the signal that gave rise to the observed interest rate r:

$$s(r) = 1/(\pi (1-\kappa)) - (1 + r_f)\ /(\pi (1-\kappa)(1+r)) - \omega_m (1-\pi)/\pi. \tag{A.6}$$

The true default probability of loans with interest rate r is the true default probability given signal s:

$$E[W|r] = E[W|s(r)] = \pi s(r) + (1-\pi)\ \omega_m. \tag{A.7}$$

Substituting equation (A.6) into equation (A.7) yields:

$$E[W\,|\,r] = \frac{1}{1-\kappa} - \frac{1+r_f}{1-\kappa}(1+r)^{-1},$$

which is identical to equation (A.2). This formalizes the intuition that for rational lenders the relationship between $E[W|r]$ and $(1+r)^{-1}$ does not depend on π, the precision of the signal that lenders receive about a loan's default probability.

Next, we allow for two departures from rationality, and examine how these departures affect the relationship between $E[W|r]$ and $(1+r)^{-1}$. First, we allow lenders to misperceive the strength of the signal they receive. Second, we allow lenders to set interest rates for reasons unrelated to the expected return (i.e., include pure noise in the interest rates chosen). One can think of lenders misperceiving the strength of the signal as lenders only perceiving a fraction of the signal (and not realizing that the unobserved fraction is perfectly correlated with the observed fraction). We think of the signal as the information content in the listing and assume that lenders only observe a fraction τ of the signal. Now, the lender's *perceived* default expectation conditional on receiving signal s is:

$$E_L[W|s] = \pi\tau s + (1-\pi\tau)\ \omega_m, \tag{A.8}$$

where ω_m is the unconditional mean default probability and the subscript L on the expectation operator denotes that the expectation is formed by the lenders. As before (see equation A.3), the true default expectation for a loan with signal s is:

$E[W | s] = \pi s + (1-\pi)\ \omega_m$.

To derive the interest rate as a function of the observed signal, we take the expectation in the no-arbitrage condition with respect to signal s but take into account that lenders only base their inference on the fraction τ of the signal:

$$(1 + r_f) = E_L[(1+r)(1-(1-\kappa)W) | s]. \qquad \text{(A.9)}$$

Substituting (A.8) into (A.9) and solving for r yields the interest rate as a function of the signal:

$$r(s) = [1/(1 + r_f) - (\pi\tau s + (1-\pi\tau)\ \omega_m)(1-\kappa)/(1 + r_f)]^{-1} - 1. \qquad \text{(A.10)}$$

We invert equation (A.10) to find the signal that gave rise to the observed interest rate r:

$$s(r) = 1/(\pi\tau (1-\kappa)) - (1 + r_f) / (\pi\tau (1-\kappa)(1+r)) - \omega_m (1-\pi\tau)/(\pi\tau). \qquad \text{(A.11)}$$

The true default probability of loans with interest rate r is the true default probability given signal s:

$$E[W | r] = E[W | s(r)] = \pi s(r) + (1-\pi)\ \omega_m. \qquad \text{(A.12)}$$

Substituting (A.11) into (A.12) yields:

$$E[W | r] = \frac{1}{(1-\kappa)\tau} - \frac{1-\tau}{\tau}\omega_m - \frac{1+r_f}{(1-\kappa)\tau}(1+r)^{-1}. \qquad \text{(A.13)}$$

Equation (A.13) implies that, as τ becomes smaller, the slope of default realizations with respect to $(1+r)^{-1}$ becomes steeper, and the relationship becomes infinitely steep (i.e., undefined) for $\tau = 0$. The intuition is that when lenders underreact to the signal ($\tau < 1$), they reduce the variance of the interest rates they set, so the data points on the *X*-axis get compressed. However, the true default probabilities do not get compressed since they do not depend on the weight that the lender places on the signal. With data points on the *X*-axis getting compressed, but those on the *Y*-axis unaffected, the relationship becomes steeper.

Next, we allow lenders to use information that is not related to the signal to set interest rates. Statistically, one can think of this as noise, which might be economically interpreted as motivations for lending unrelated to returns (altruism, taste-based discrimination, etc.). We model this unrelated information by the error term η, and write:

$$(1+r)^{-1} = (1+r^*)^{-1} + \eta, \qquad \text{(A.14)}$$

where $(1+r)^{-1}$ is the actual interest rate charged and $(1+r^*)^{-1}$ is the interest rate lenders would have charged if they behaved purely to maximize expected returns (though they may still misperceive the strength of the signal). Thus, r^* is given by equation (A.10) above, so that:

$$(1+r^*)^{-1} = 1/(1 + r_f) - (\pi\tau s + (1-\pi\tau)\ \omega_m)(1-\kappa)/(1 + r_f). \qquad \text{(A.15)}$$

Let the error term be uncorrelated with the "underlying" interest rate $(1+r^*)^{-1}$, and let the variance of η be denoted by σ_η^2. The variance in $(1+r^*)^{-1}$ can be found by noting that the only stochastic term in (A.15) is *s*, and that the variance of *s* is equal to the variance of the true default probability, denoted by σ_ω^2. The variance of $(1+r^*)^{-1}$ is thus equal to:

$$\sigma^2_{(1+r^*)^{-1}} = \left(\pi\tau(1-\kappa)/(1+r_f)\right)^2 \sigma^2_\omega . \tag{A.16}$$

The relationship between true default probability and $(1+r^*)^{-1}$ is still given by (A.13). If we regress the observed ex-post default rates on the actual interest rate $(1+r)^{-1}$, which is a noisy measure of the underlying interest rate $(1+r^*)^{-1}$, the coefficient on the actual interest rate will be attenuated, with degree of attenuation given by the standard formula for attenuation bias from classical measurement error. The coefficient of a regression of default probabilities on $(1+r)^{-1}$ will therefore be:

$$-\frac{1+r_f}{(1-\kappa)\tau}\bullet\left[1-\frac{\sigma^2_\eta}{\sigma^2_{(1+r^*)^{-1}}+\sigma^2_\eta}\right] = -\frac{1+r_f}{(1-\kappa)\tau}\bullet\left[1-\frac{\sigma^2_\eta}{\left(\pi\tau(1-\kappa)/(1+r_f)\right)^2\sigma^2_\omega+\sigma^2_\eta}\right]. \tag{A.17}$$

Expression (A.17) shows that if lenders base their interest rates partly on information unrelated to default probabilities ("noise"), the relationship between default realizations and $(1+r)^{-1}$ will become flatter. The intuition is that the noise spreads out the data points on the *X*-axis, leading to a flatter regression line. Thus, the two departures from rationality – (i) incomplete inference from the signal and (ii) noise in setting the interest rate – have opposite predictions on the slope of the relationship between default realizations and $(1+r)^{-1}$. Since we have no way of estimating σ_η^2 and σ_ω^2, we cannot use our estimate of the slope to estimate τ, the degree of inference that lenders make from the signals they observe. This is why we focus on the R^2 and AUC measures instead.

Online Appendix B: Methodology to estimate inference along the credit-score dimension

The benefit of the stylized setup shown in Figure 2 and the corresponding regression is their simplicity. However, if the true credit score were observable, the underlying relationship between interest rate and exact credit score could very well be non-linear. Moreover, credit categories are not all of equal size. Online Appendix Figure C.1 depicts this more realistic situation. The dashed blue line shows a possible underlying relationship between interest rate and exact credit score for the hypothetical scenario in which exact credit score is observable by lenders. This relationship is now allowed to be non-linear. As a result of this non-linearity, the slope of the observed relationship between market interest rate and credit score need not be the same within each credit category, and the jump in market interest rate at the category borders may vary. The solid red line depicts the estimated relationship between market interest rate and exact credit score. This line falls by β_k within category k and falls by α_k at the border between category k-1 and category k.

To determine the amount of inference, we first calculate the total fall in interest rate over each credit category. To do so, we need to decide what part of the jump of size α_k at the border between category k-1 and category k can be attributed to category k-1 and what part to category k. It appears most natural to attribute this jump proportionally to the size of each category, but results are similar when we attribute it evenly across the two bordering categories. Let λ_k denote the size of category k-1 as a fraction of the combined size of categories k-1 and k. Then, the part of the drop in interest rate at the border of categories k-1 and k that is attributed to category k is equal to $(1-\lambda_k)\alpha_k$. Similarly, the part of the drop at the next category border that is attributed to category k is $\lambda_{k+1}\alpha_{k+1}$. Since the interest rate falls by β_k *within* category k, the total drop in interest associated with category k is $\delta_k=(1-\lambda_k)\alpha_k+\lambda_{k+1}\alpha_{k+1}+\beta_k$.[2] The fraction of information inferred within this category, γ_k, can then be calculated as β_k/δ_k.

To estimate these parameters, we regress the interest rate on a spline in the exact credit score and cumulative dummies for the credit-score categories:

$$InterestRate_i = \mu + \sum_{k=2}^{N} \alpha_k I_k^{Cum}(CreditScore_i) + \sum_{k=1}^{N} \beta_k FracGap_k(CreditScore_i) + \varepsilon_i \,, \tag{B.1}$$

where $InterestRate_i$ is the interest rate charged on loan i, $CreditScore_i$ is the exact credit score of the borrower of loan i, $I_k^{Cum}(CreditScore_i)$ are cumulative credit-score dummies, and $FracGap_k$ is a variable that increases linearly with exact credit score within credit category k and that is constant everywhere else. The coefficient α_k measures the jump in interest rate at the credit-score boundary between credit categories k-1 and k, the coefficient β_k measures the change in interest rate within category k, and ε_i is the error term. Formally, we define $I_k^{Cum}(CreditScore_i)$ as an indicator variable that equals one if borrower i is in credit category k or higher:

[2] By definition, we can neither estimate a jump at the lower border of the bottom credit category nor at the upper border of the top credit category. When calculating the gammas for the first (bottom) and seventh (top) category, we assume that jumps at the lower and upper borders are of equal size: We assume that $(1-\lambda_1)\alpha_1$ equals $\lambda_2\alpha_2$ and that $\lambda_8\alpha_8$ equals $(1-\lambda_7)\alpha_7$.

$$I_k^{Cum}(CreditScore_i) = \begin{cases} 0 & \text{if} \quad CreditScore_i < c_k \\ 1 & \text{if} \quad CreditScore_i \geq c_k \end{cases}, \tag{B.2}$$

where c_k is the credit score that forms the boundary between categories k-1 and k. Formally, $FracGap_k(CreditScore_i)$ is defined as:

$$FracGap_k(CreditScore_i) = \begin{cases} 0 & \text{if} \quad CreditScore_i \leq c_k \\ \dfrac{CreditScore_i - c_k}{c_{k+1} - c_k} & \text{if} \quad c_k < CreditScore_i \leq c_{k+1} \\ 1 & \text{if} \quad c_{k+1} < CreditScore_i \end{cases}, \tag{B.3}$$

Thus, $FracGap_k$ increases linearly from 0 to 1 as we move from the lowest to the highest credit score within category k. Further, $FracGap_k$ is 0 for values below c_k and equals 1 for all credit scores above c_{k+1}.

When we estimate equation (B.1), the test $\beta_k = 0$ tests the hypothesis that lenders are not able to infer variation in creditworthiness within category k (along the dimension measured by exact credit score) from all the information provided in the listing. Since the estimates of the β_k may be relatively imprecise, we also test the joint hypothesis that all β_k are equal to zero. Because the coefficients α_k measure the jumps in interest rate at the credit-score boundaries, we can reject the hypothesis that lenders perfectly infer creditworthiness (along the dimension measured by exact credit score) from the information on the listing if these αs are statistically significant.

Because we estimate the γ parameters separately for each credit category, they are each based on relatively few observations. As a result, the parameters may not be estimated very precisely for particular categories, even if they are jointly significant. We therefore also present a combined γ estimate, which is the weighted average across credit-score categories of γ_k, where the weights are the precision with which the parameter is estimated in each category.

When we estimate equation (B.1), we hope to recover the effect of the listing characteristics on the interest rate that lenders require to compensate them for the perceived credit risk of that listing. If this interest rate exceeds the maximum interest rate that the borrower is willing to pay (as specified by the variable *borrower maximum rate*), the listing will not be funded, and we consequently do not observe the interest rate that lenders require. Thus, our observations of the interest rate are censored at the borrower maximum rate.[3] This censoring problem would bias our estimates of inference if we were to estimate equation (B.1) using ordinary least squares. Instead, we estimate equation (B.1) as a censored regression with the censoring occurring at the borrower maximum rate specified by each listing. A censored regression, which is a generalization of the Tobit model, rests on the implicit assumption that listings that were not funded would have been funded at some interest rate larger than the observed borrower maximum rate. If the error term has a homoscedastic and normal distribution, estimates from a censored regression will yield consistent estimates of the parameters determining the interest rates that lenders require to fund a listing.

To estimate the highest gamma that could have been achieved if lenders had perfectly used all hard and coded soft information available in the listing sheet, we first regress default realizations on a very flexible functional form of all standard financial variables and soft/nonstandard variables.[4]

---

[3] State usury laws limit the maximum interest rate that borrowers may set for loans (most states allow a maximum interest rate of 36%). Thus, when state usury caps censor the market interest rate, the usury cap censors at the borrower maximum rate.

[4] Specifically, we use a flexible set of 215 controls for standard financial and nonstandard variables, as described in Table

We use this regression to predict a default probability for each loan. Because these predicted default probabilities are based on information that was observable to the lenders, lenders could also have made these predictions themselves (if they had used all coded listing content optimally). Next, we rerun our baseline specification (equation B.1) using the predicted default probability as the dependent variable instead of the interest rate. The gamma from this baseline specification with predicted default as the dependent variable tells us what fraction of predictable default along the credit score dimension occurs *within* credit categories. Hence, we consider it a benchmark for our baseline estimates because it tells us what gamma the lenders could have achieved had they adjusted their interest rate for the default risk that they could have predicted based on coded listing content.

A caveat to our methodology is that we measure inference of credit score *within* credit categories – our measurements necessarily take the size of credit categories as given. Prosper exogenously sets credit categories to represent 40-point increments in credit score. While it is not obvious whether this poses a systematic concern, it is possible that our estimates could be different if, for example, Prosper created a different credit category for every ten points in credit score instead of for every 40 points. For example, for very narrow credit-score bins, lenders may exert less effort to infer differences in credit score within bins and hence (optimally) show lower learning inference. Thus, in a sense, the credit category needs to be large enough for inferring quality *within* it to be economically meaningful. In our case, there is substantial variation in credit quality within credit categories. A forty-point change in credit score represents a large range in creditworthiness, as evidenced by the average 400-basis-point decline per credit category in the mean interest rate offered by lenders.

Decomposing Inference by Source of Information

The inference parameter γ measures the contribution of all sources of information on the Prosper website, whether or not this information can be coded as a quantitative variable. To measure the contributions of various information sources, we add to regression (B.1) controls for all the coded content of the listing:

$$InterestRate_i = \mu + \sum_{k=2}^{N} \alpha_k I_k^{Cum}(CreditScore_i) + \sum_{k=1}^{N} \beta_k^{Resid} FracGap_k(CreditScore_i) + \sum_{m=1}^{M} x_i^m \varphi^m + \varepsilon_i, \qquad \text{(B.4)}$$

where x_i^m denotes the m^{th} variable for the coded listing content of borrower i, and φ^m denotes the corresponding regression coefficient.[5] In regression (B.4), the fitted interest rate can change with credit score within a credit category for two reasons. First, even after controlling for all the observable characteristics, there still may be a residual correlation between exact credit score and interest rate within a credit category due to inference from uncoded listing content. Since we measure exact credit score within credit categories by *FracGap*, this residual correlation is measured by β_k^{Resid}. Second, the fitted interest rate may vary within a credit category because (i) listings with higher values of *FracGap* may have different observable characteristics and (ii) the interest rate responds to these characteristics. We measure component (i) – the degree to which observable

1, which are further interacted with the seven credit category dummies.

[5] In all specifications, we define the x variables to be specific within credit categories, which means that we estimate the φ coefficients for the control variables separately by credit category. We correct the α coefficients for any jumps in the interest rate at credit-category boundaries that are absorbed by the interactions of x and the credit categories or for jumps in the x variables themselves. This correction ensures that the α coefficients fully capture the jumps in the interest rate at the category boundaries.

characteristic x^m varies with *FracGap* – by running a regression of the observations of x^m within category k on $FracGap_k$ and a constant term. We denote the coefficient on $FracGap_k$ in this bivariate regression by θ_k^m.We measure component (ii) – the degree to which the interest rate responds to characteristic x^m – by the regression coefficient φ^m. The total contribution of variable x^m to the relationship between *FracGap* and interest rate within category k is given by the product of these two components: $\beta_k^m \equiv \theta_k^m \varphi^m$.

We decompose our original estimate β_k from the regression without the controls for coded listing characteristics (regression B.1) as follows:[6]

$$\beta_k = \beta_k^{Resid} + \sum_{m=1}^{M} \theta_k^m \varphi^m \equiv \beta_k^{Resid} + \sum_{m=1}^{M} \beta_k^m . \tag{B.5}$$

In equation (B.5), $\sum_{m=1}^{M} \beta_k^m$ is the part of the within-category drop in interest rates that can be attributed to coded information, while the remainder is explained by uncoded information. Thus, rather than attempting to code the qualitative information (quantification of which, by definition, will be highly imperfect), we infer its information content from β_k^{Resid}, which measures the extent to which the interest rate varies with exact credit score within credit-score categories after controlling for all coded information. To ensure that β_k^{Resid} reflects uncoded information, rather than omitted higher-order terms of the x variables, we include all x variables as quadratics and interact them with credit-category indicators.[7] Instead of reporting each single β_k^m, we report a sum of the βs that correspond to standard financial variables and a sum of the βs that correspond to soft/nonstandard variables. We also include β_k^{Resid}, which measures the contribution of uncoded information, with the soft/nonstandard variables. Finally, the corresponding inference parameters, γ_k^m, are calculated by dividing each type of β_k by δ_k.

We should note that this decomposition is accurate, provided that listing characteristic x^m affects interest rates only through the aspect of creditworthiness captured by credit score. Alternately, φ^m may capture an effect of x^m on the interest rate that is mediated both through the credit-score dimension and another dimension of creditworthiness. In that case, we would ascribe less (more) inference to x^m if it has a similar (opposite) impact on this other dimension of creditworthiness (compared to the credit-score dimension).

---

[6] This is an application of the standard omitted variable bias formula. The omitted variable bias formula holds by construction if the equation is estimated by OLS. However, because we estimate our model as a censored regression, the omitted variable bias decomposition holds only in expectation. As a result, our decomposition will not add up exactly.

[7] In addition, we include dummy variables for each of the following variables taking on a value of zero: Number of Current Delinquencies, Number of Delinquencies in Last 7 Years, Number of Public Record Requests in Last 10 Years, Number of Public Records in Last 12 Months, and Revolving Credit Balance. Amount Delinquent and Revolving Credit Balance are not introduced as quadratics, but as logs with dummies for values equal to zero and values less than or equal to 100. Each of these variables (including dummies and higher-order terms) is interacted with a full set of credit-category indicators.

Appendix C, Table C.1: Decomposing Inference, Part I (Standard Financial Variables)

	(1) All Credit Categories			(2) Low Credit Categories (HR - C)			(3) High Credit Categories (B - AA)			(4) Low = High p-value
Standard Financial Variables										
No. of Current Delinquencies	0.079	(0.006)	***	0.110	(0.010)	***	0.045	(0.007)	***	0.000
No. of Credit Inquiries, Last 6 Months	0.054	(0.003)	***	0.073	(0.004)	***	0.034	(0.003)	***	0.000
Amount Delinquent	0.051	(0.006)	***	0.085	(0.010)	***	0.015	(0.006)	***	0.000
Debt-to-Income Ratio	0.048	(0.007)	***	0.001	(0.008)		0.099	(0.011)	***	0.000
Amount Requested	-0.005	(0.005)		-0.124	(0.006)	***	0.122	(0.009)	***	0.000
No. of Delinquencies, Last 7 Years	0.033	(0.004)	***	0.043	(0.006)	***	0.023	(0.005)	***	0.006
No. of Public Records, Last 10 Years	0.023	(0.002)	***	0.018	(0.004)	***	0.028	(0.003)	***	0.056
Total No. of Credit Lines	-0.004	(0.005)		-0.008	(0.009)		0.001	(0.005)		0.391
Bank Card Utilization Ratio	-0.003	(0.011)		0.008	(0.006)		-0.015	(0.021)		0.290
No. of Public Records, Last 12 Months	0.000	(0.002)		-0.001	(0.002)		0.000	(0.003)		0.896
No. of Current Credit Lines	0.004	(0.008)		0.006	(0.015)		0.002	(0.006)		0.807
No. of Open Credit Lines	-0.002	(0.008)		-0.001	(0.014)		-0.002	(0.006)		0.945
Revolving Credit Balance	-0.011	(0.007)		-0.025	(0.010)	***	0.005	(0.010)		0.028
Homeownership Dummy	0.024	(0.006)	***	0.011	(0.005)	**	0.039	(0.010)	***	0.013
Credit History Age	0.007	(0.005)		0.010	(0.007)		0.004	(0.007)		0.558
State of Residency (52 Dummies)	-0.013	(0.005)	***	-0.024	(0.007)	***	-0.002	(0.006)		0.024
Employment Status (5 Dummies)	0.002	(0.002)		0.007	(0.004)	*	-0.004	(0.001)	**	0.009
Length of Current Employment Status	-0.003	(0.001)	**	-0.005	(0.002)	**	-0.001	(0.001)		0.059
Personal Annual Income (7 Dummies)	0.014	(0.005)	***	0.012	(0.006)	**	0.016	(0.009)	*	0.711
Borrower Occupation (62 Dummies)	0.011	(0.006)	**	0.011	(0.008)		0.011	(0.007)		0.990
Missing Data (2 Dummies)	0.001	(0.002)		0.003	(0.002)		0.000	(0.003)		0.464

This table shows the decomposition of our estimate of gamma presented in Table 6, Column (2). The decomposition results are divided into standard financial variables, presented here, and soft/nonstandard variables, presented in the next page. The decomposition is based upon the baseline censored normal specification with the addition of 216 control variables, each interacted with seven credit category dummies, such that the coefficient on each control variable is allowed to vary by credit category. All controls except for dummy variables are entered as quadratics. *Amount delinquent* and *revolving credit balance* are introduced as logs with dummies for values equal to zero and values less than or equal to 100. *Missing Data* consists of two dummies equal to one when subsets of the standard financial variables are missing in the data (observations with missing standard financial variables account for less than one percent of our sample). Column (1) presents results for the entire sample, while the next two columns, (2)-(3), present the combined gamma separately for the lower credit categories (C, D, E, and HR) and the higher credit categories (AA, A, and B). Column (4) presents the p-value from a test of whether the estimates for the lower and higher credit categories are equal. Standard errors are allowed to be clustered by borrower (some borrowers apply for more than one loan) and are in brackets with * significant at 10%; ** significant at 5%; and *** significant at 1%.

Appendix C, Table C.1: Decomposing Inference, Part II (Nonstandard Variables)

	(1) All Credit Categories			(2) Low Credit Categories (HR - C)			(3) High Credit Categories (B - AA)			(4) Low = High p-value
Non-Standard Variables										
Borrower Maximum Interest Rate	0.064	(0.004)	***	0.083	(0.005)	***	0.043	(0.007)	***	0.000
Listing Category (8 Dummies)	-0.026	(0.003)	***	-0.048	(0.005)	***	-0.002	(0.005)		0.000
Member of Group Dummy	-0.016	(0.002)	***	-0.028	(0.004)	***	-0.003	(0.001)	***	0.000
Group Leader Reward Rate (9 Dummies)	-0.015	(0.002)	***	-0.028	(0.004)	***	-0.002	(0.002)		0.000
Duration of Loan Listing (4 Dummies)	-0.011	(0.002)	***	-0.009	(0.003)	***	-0.012	(0.003)	***	0.447
Bank Draft Annual Fee Dummy	0.000	(0.001)		0.000	(0.001)		0.000	(0.001)		0.924
Borrower Lists City Dummy	-0.001	(0.001)		-0.001	(0.002)		0.000	(0.000)		0.623
Borrower Provides Image Dummy	-0.002	(0.001)	**	-0.004	(0.001)	***	0.000	(0.001)		0.044
HTML Character No.	0.000	(0.002)		-0.001	(0.002)		0.001	(0.002)		0.542
Text Character No.	-0.005	(0.002)	***	-0.006	(0.004)		-0.005	(0.001)	***	0.808
Average Word Length	0.002	(0.001)		0.004	(0.003)		-0.001	(0.001)		0.075
Average Sentence Length	-0.003	(0.001)	**	-0.007	(0.002)	***	0.002	(0.001)	**	0.001
No. of Numerics	-0.003	(0.004)		0.000	(0.003)		-0.006	(0.008)		0.510
Percent Misspelled	-0.001	(0.001)		-0.001	(0.002)		0.000	(0.001)		0.502
No. of Dollar Signs	-0.003	(0.004)		-0.003	(0.003)		-0.003	(0.008)		0.983
Percent of Listing as Signs	0.003	(0.002)	**	0.004	(0.003)		0.002	(0.001)		0.570
No. of Characters in Listing Title	-0.001	(0.001)		-0.002	(0.001)		0.000	(0.001)		0.292
No. of Friend Endorsements	-0.007	(0.002)	***	-0.016	(0.003)	***	0.003	(0.003)		0.000
Uncoded Listing Content	0.040	(0.032)		0.096	(0.045)	**	-0.020	(0.044)		0.066

This table shows the decomposition of our estimate of gamma presented in Table 6, Column (2). The decomposition results are divided into standard financial variables, presented in the previous page, and soft/nonstandard variables, presented here. The decomposition is based upon the baseline censored normal specification with the addition of 216 control variables, each interacted with seven credit category dummies, such that the coefficient on each control variable is allowed to vary by credit category. We control for coded standard and non-standard variables as quadratics, with Amount Delinquent and Revolving Credit Balance measured in log form. We also include dummy variables for each of the following variables taking on a value of zero: Number of Current Delinquencies, Number of Delinquencies in Last 7 Years, Number of Public Record Requests in Last 10 Years, Number of Public Records in Last 12 Months, Revolving Credit Balance, and Amount Delinquent. We further include dummy variables for Amount Delinquent and Revolving Credit Balance less than $USD100. Missing Data consists of two dummies equal to one when subsets of the standard financial variables are missing in the data (observations with missing standard financial variables account for less than one percent of our sample). Column (1) presents results for the entire sample, while the next two columns, (2)-(3), present the combined gamma separately for the lower credit categories (C, D, E, and HR) and the higher credit categories (AA, A, and B). Column (4) presents the p-value from a test of whether the estimates for the lower and higher credit categories are equal. Standard errors are allowed to be clustered by borrower (some borrowers apply for more than one loan) and are in brackets with * significant at 10%; ** significant at 5%; and *** significant at 1%.

Table C.2: Interest Rates and Loan Performance, Double Clustering by Time and Borrower

Panel A: OLS - Do Interest Rates Predict Loan Performance?	Default - 3 or more months late			Fraction Repaid		
1/(1 + Interest Rate)	-1.525	***	(0.116)	1.173	***	(0.085)
N	17212			17212		
Adjusted R^2	0.0814			0.0914		

Panel B: OLS - Do Credit Scores Predict Loan Performance?	Default - 3 or more months late			Fraction Repaid		
Exact Credit Score/100	-0.129	***	(0.009)	0.096	***	(0.007)
N	17212			17212		
Adjusted R^2	0.0432			0.0456		
Adjusted R^2 using 7-part spline in credit score	0.0442			0.0476		

Panel C: IV - Do Interest Rates Causally Affect Loan Performance?	Default - 3 or more months late			Fraction Repaid		
1/(1 + Interest Rate)	0.166		(0.415)	0.061		(0.298)
N	17212			17212		
First Stage F-stat	58.45			58.45		

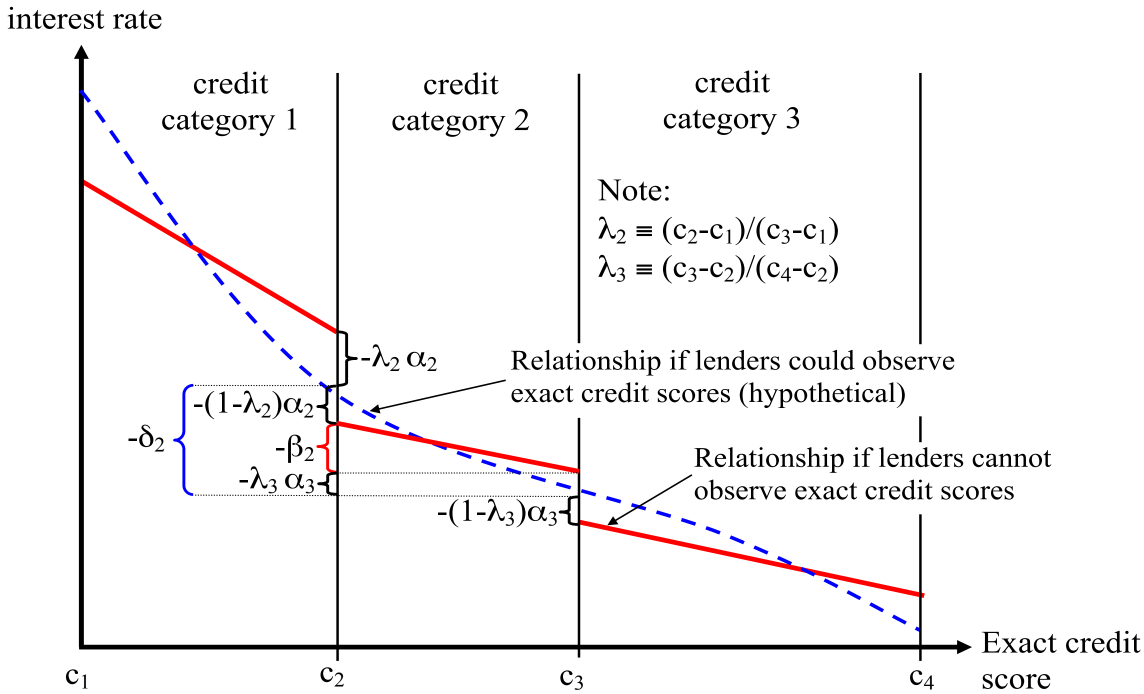
This table is identical to Table 2, except that the standard errors are double clustered by time (year*month) and by borrower. For further notes, see notes to Table 2.

Table C.3: Interest Rates and Loan Performance, Double Clustering by Lender and Borrower

Panel A: OLS - Do Interest Rates Predict Loan Performance?	**Default - 3 or more months late**			**Fraction Repaid**		
1/(1 + Interest Rate)	-1.525	***	(0.047)	1.173	***	(0.035)
N	17212			17212		
Adjusted R^2	0.0814			0.0914		
Panel B: OLS - Do Credit Scores Predict Loan Performance?	**Default - 3 or more months late**			**Fraction Repaid**		
Exact Credit Score/100	-0.129	***	(0.006)	0.096	***	(0.004)
N	17212			17212		
Adjusted R^2	0.0432			0.0456		
Adjusted R^2 using 7-part spline in credit score	0.0442			0.0476		
Panel C: IV - Do Interest Rates Causally Affect Loan Performance?	**Default - 3 or more months late**			**Fraction Repaid**		
1/(1 + Interest Rate)	0.166		(0.415)	0.061		(0.298)
N	17212			17212		
First Stage F-stat	58.45			58.45		

This table is identical to Table 2, except that the standard errors are double clustered by lender and by borrower. For further notes, see notes to Table 2.

Appendix C, Figure C.1: Relationship between Interest Rate and Credit Score



This figure shows a more realistic hypothesized relationship between a borrower's credit score and the market interest rate on her (funded) loan.

Appendix D: Sample Listing

PROSPER® Let's bank on each other™

Home | Get a Loan | Bid on Loans | Community | My Account | Help

Search Listings | Portfolio Plans | Advanced Search | About Lending | Rates | Performance | Watch List

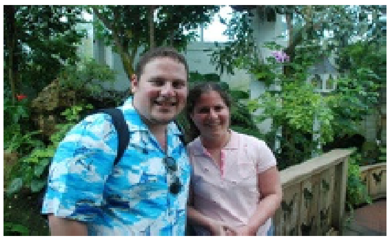
help me pay off credit cards and propose to my girlfriend

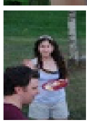
(Listing #208364)

« Back to search results

LISTING SUMMARY

Help





$8,081.00 @ 8.90%

Bid down from 13.99%

Bid Now

(Bidding has ended)

Funding:	100% funded
Bids:	321 bids Ended Listing became a loan
Borrower APR:	9.59%
Mo. payment:	$256.60 (3y loan)

Watch | Email | Report this listing

BORROWER INFO

Help

hs4g2

CANTON, MA

Members and Friends of the Boston Area College Community

(33)

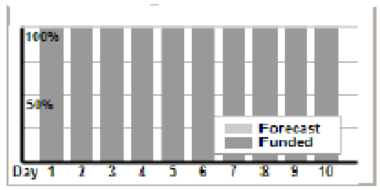
0 friend bids

0 questions & answers

0 friends, 0 verified

1 loan total, 1 active

FORECAST | COMPARE | Help



CREDIT PROFILE

Help

A credit grade — Homeownership not verified

Now delinquent:	0	**First credit line:**	Mar-2001
Amount delinquent:	$0	**Current / open credit lines:**	3 / 3
Delinquencies in last 7y:	0	**Total credit lines:**	4
Public records last 12m / 10y:	0 / 0	**Revolving credit balance:**	$3,531
Inquiries last 6m:	1	**Bankcard utilization:**	29%

Credit and homeownership information provided by Experian.

10% debt to income ratio

Employment status:	Full-time employee
Length of status:	1y 0m
Stated income:	$25,000-$49,999
Occupation:	Computer Programmer

Employment and income provided by borrower.

DESCRIPTION

Purpose of loan:

I'm using this loan to pay off my $3,531 credit card bill currently at 14% at a lower interest rate and to buy my girlfriend, Jennifer, an engagement ring costing approx. $4,550. Up until about 3 months ago, I would revolve all my purchases through my credit card and I ended up letting it get slightly away from me. As a result, I've devised a plan to pay off as much debt as possible per month (currently, I pay $550 to my credit card company) and live on a necessity only budget. The next phase of my plan after eliminating my credit card debt was to immediately go back into almost as much debt as I have now to buy Jenn a ring. Then along came prosper. With your help, I'll be able to ask Jenn to marry me sooner than expected and maybe not even be in debt when I do it!

My financial situation:

Currently, I work as a software engineer in Wellesley, MA. I make a pretty good living and enjoy what I do. The people I work with like and respect me and I feel my job is very secure and also portable (i.e. I can work from anywhere with an internet connection) should I need to move (Jenn is in her 4th year of med school and is looking at residencies). Below, you can see my monthly expenses which will be going down come May/June since Jenn and I will be moving in together. I invest in the stock market and I am also using Prosper on the lender's side. I have a little bit of cash set aside for a rainy day and a bit more available (though not as quickly attainable) in case of a financial hurricane. I also put away 12% of my gross pay into a 401k which my company contributes to with profit sharing.

Appendix: Sample Listing - Continued

Monthly net income: $ 2084

Monthly expenses: $ 1705
Housing: $ 535
Insurance: $ 200
Car expenses: $ 125
Utilities: $ 40
Phone, cable, internet: $ 85
Food, entertainment: $ 400
Clothing, household expenses $ 50
Credit cards and other loans: being paid with this loan
Other expenses: $ 0
Prosper Loan: $270

FRIENDS AND FAMILY WINNING BIDS

Help

This member has no winning bids from friends and family.

QUESTIONS & ANSWERS

This borrower has not publicly answered any questions from lenders.

BID HISTORY

Legend: = In group = Friend ✓ = Winning ✓ = Partially winning ✗ = Outbid

Help

Bidder / Relationship	Rate	Amount Bid	Winning	Status ▲	Bid Date (PT)
wolfpac79	8.90%	$50.00	$50.00	✓	Oct-09-2007 8:26 AM
user13	8.90%	$50.00	$50.00	✓	Oct-09-2007 8:13 AM
JDLanier	8.90%	$50.00	$50.00	✓	Oct-09-2007 8:11 AM
steamboatgal	8.90%	$100.00	$100.00	✓	Oct-09-2007 8:00 AM
lender1853	8.90%	$100.00	$100.00	✓	Oct-09-2007 7:56 AM
Porsche2	8.90%	$50.00	$50.00	✓	Oct-09-2007 7:52 AM
mmmoney	8.90%	$100.00	$100.00	✓	Oct-09-2007 7:47 AM
mmmoney	8.90%	$100.00	$100.00	✓	Oct-09-2007 7:40 AM
universe	8.90%	$75.00	$75.00	✓	Oct-09-2007 7:35 AM
swissbanker	8.90%	$100.00	$100.00	✓	Oct-09-2007 7:28 AM
OGS_Capital	8.90%	$51.42	$51.42	✓	Oct-09-2007 7:24 AM
moose_spencer	8.90%	$50.00	$50.00	✓	Oct-09-2007 7:16 AM
Orphan2007	8.90%	$50.00	$50.00	✓	Oct-09-2007 6:54 AM
wkk	8.90%	$50.00	$50.00	✓	Oct-09-2007 6:50 AM
LoanChimp	8.90%	$100.00	$100.00	✓	Oct-09-2007 6:29 AM
Gromila18	8.90%	$200.00	$200.00	✓	Oct-09-2007 6:01 AM
Badger1	8.90%	$50.00	$50.00	✓	Oct-09-2007 5:53 AM
Goodthings2you	8.90%	$50.00	$50.00	✓	Oct-09-2007 5:26 AM
stevedavis444	8.90%	$50.00	$50.00	✓	Oct-09-2007 5:13 AM
Deal_Flow	8.90%	$50.00	$50.00	✓	Oct-09-2007 5:06 AM
Curlingman	8.90%	$50.00	$50.00	✓	Oct-09-2007 4:59 AM
5Star	8.90%	$50.00	$50.00	✓	Oct-09-2007 4:49 AM