

Facing Default?*

Pranav Garg

Yale

Marius Guenzel

Wharton

Shimon Kogan

Wharton/Reichman

Marina Niessner

Indiana

Kelly Shue

Yale

December 6, 2025

Preliminary. Do not circulate without permission.

Abstract

We study whether AI-extracted facial features from borrowers' photos can serve as a scalable proxy for “soft” information missing from traditional credit models, such as conscientiousness, patience, and self-control. These traits influence financial behavior but are rarely captured in administrative data. Linking LinkedIn photos and employment and education records to voter registration and Experian data for over one million U.S. borrowers, we find that facial embeddings add significant predictive power for default risk beyond standard observables such as credit scores, gender, and race. The incremental value is largest for younger, lower-income, and thin-file borrowers, where traditional credit scoring technology is least informative. A separate model mapping facial images to perceived Big Five personality traits reveals personality as one mechanism through which images proxy for soft information. These results suggest that facial embeddings capture stable behavioral traits absent from standard credit data, as well as perceived attributes that may influence how individuals are treated by others. While such models offer new insight into the role of personality and soft information in credit markets, their use in screening raises important concerns about fairness, privacy, and autonomy.

Keywords: Credit risk, soft information, AI and machine learning, behavioral traits, perceived personality

*This draft has benefited from comments by Tony Cookson as well as presentations at the 2025 Boulder Summer Conference on Consumer Financial Decision Making and the University of Maryland. We thank Dilip Francies for providing invaluable research assistance. This study was approved by the Institutional Review Board at Yale University under protocol number 2000036766.

Contact: Pranav Garg (pranav.garg@yale.edu), Marius Guenzel (mguenzel@wharton.upenn.edu), Shimon Kogan: (skogan@wharton.upenn.edu), Marina Niessner (mniessne@iu.edu), Kelly Shue (kelly.shue@yale.edu).

1. INTRODUCTION

Who gets credit, on what terms, and why, are first-order questions. Classic models of credit markets formalize them by framing lending as a problem of information (Jaffee and Russell, 1976, Stiglitz and Weiss, 1981, Sharpe, 1990). Testing the seminal theories of credit market behavior and outcomes requires empirically identifying the private *types* of economic agents that underpin such models (e.g., “honest” versus “dishonest,” “high” versus “low” quality, and “safe” versus “risky”). Modern credit scoring draws on extensive “hard” information to estimate borrower types, but widely used credit scoring models remain uneven in their efficacy (Blattner and Nelson, 2021, Fuster et al., 2022, Bartlett et al., 2022). Scores are most informative for borrowers with long, thick files and least informative precisely where stakes are highest: younger, lower-income, and thin-file consumers with limited credit histories.

Beyond hard observables, a large share of default risk may be tied to “soft” information inherent to the borrower – latent traits such as conscientiousness, patience, self-control, and a propensity to follow rules – along with perceptions these traits create in others. Together, these factors may shape day-to-day financial behavior but are rarely captured in administrative data. Alternative behavioral credit data, such as digital footprints or social networks or transaction patterns, are increasingly used to proxy for these dimensions (Lin et al., 2013, Iyer et al., 2016, Berg et al., 2020), but they remain costly to collect, limited in coverage, and unevenly distributed in ways that mirror existing gaps in traditional credit files.

In this paper, we examine whether information embedded in borrowers’ facial characteristics can provide a scalable proxy for the “soft” attributes that traditional credit models overlook. We assemble a large dataset linking LinkedIn profile images and professional records to voter registration and credit bureau data for more than one million U.S. borrowers. We then train machine-learning (ML) models to extract facial embeddings from the LinkedIn photographs and evaluate their predictive power for credit scores and delinquency, both unconditionally and after controlling for a rich set of demographic, employment, and education characteristics. All performance is evaluated entirely out of sample.

Our academic evaluation of these capabilities is grounded in emerging real-world prac-

tice. In many settings, lenders have already begun integrating facial technologies into credit processes. For instance, roughly 70% of Brazilian banks now use facial recognition for client identification ([Agência Brasil, 2024](#)) and Pakistan’s central bank explicitly endorses artificial intelligence (AI)-based facial recognition for digital onboarding ([State Bank of Pakistan, 2025](#)). Similarly, in China, major institutions have piloted face-expression analysis during loan interviews to flag misrepresentation and fraud risk ([Financial Times, 2018](#), [Biometric Update, 2018](#), [Ping An, 2019](#)). Finally, use of alternative data is rising even among U.S. lenders: in 2024, nearly half of personal loans originated from non-bank digital lenders, though these models generally do not, at least not yet, incorporate facial information.

There are several non-mutually exclusive channels through which facial features may convey information about borrowers, particularly forms of “soft” information absent from standard observables. First, biological and developmental factors may jointly shape facial structure and behavioral tendencies. Second, stable personality traits may manifest through enduring facial cues (e.g., micro-expressions or habitual muscle tone). Third, expressive styles and grooming choices may correlate with financial behavior and self-discipline. Collectively, these mechanisms suggest that facial features can encode information about underlying behavioral types and the perceptions they elicit in others ([Claes et al., 2014](#), [Vukasović and Bratko, 2015](#), [Kachur et al., 2020](#), [Peterson et al., 2022](#), [Athey et al., 2022](#), [Guenzel et al., 2025](#)).

Importantly, we emphasize that the goal of our work is *not* to advocate deploying face-based credit screening or lending decisions. While facial images may offer a scalable proxy for “soft” information (particularly for thin-file borrowers early in their credit histories, for whom standard scores are noisy, slow to update, and often fail to level the playing field (e.g., [The Economist, 2016](#), [Berg et al., 2020](#))), their use raises serious concerns. Face-based predictions can be viewed as a form of classical statistical discrimination ([Phelps, 1972](#)), since they rely on aggregate correlations between immutable (or difficult-to-alter) facial features and behavioral traits, echoing the very concerns that motivated the Fair Credit Reporting Act in the late 1960s and that continue to shape today’s algorithmic-bias debates around privacy, transparency, and disparate impact ([Kleinberg et al., 2018](#)). The purpose of our analysis is to document these relationships empirically, in response to rapidly evolving credit-industry

practices, and to deepen our understanding of whether and how “soft” factors affect credit outcomes. We use machine learning as “a tool for scientific discovery” to uncover hidden patterns in large datasets and illuminate where residual information resides and why (Ludwig and Mullainathan, 2024).

To study the extent to which AI-extracted facial features from borrowers’ photos proxy for “soft” information, we construct a novel person-level representative sample covering more than one million U.S. borrowers. Specifically, we link employment histories from Revelio Labs (which contains data on LinkedIn profiles including profile photos) to validated demographics from L2 voter files, and further to credit outcomes from Experian. The analysis focuses on the 2019 Experian vintage, augmented with each individual’s 2017 snapshot to construct lagged covariates. Our final dataset contains granular information on standard borrower characteristics, credit outcomes such as Vantage scores and delinquency rates, as well as individual education and labor histories.¹

We then train a ML pipeline that extracts facial features from individuals’ LinkedIn profile photos in order to predict credit outcomes. In particular, we extract a 512-dimensional facial embedding vector using the ArcFace deep-learning model (Deng et al., 2019). ArcFace, trained on millions of labeled face images, maps each face into a standardized numerical representation that captures stable, person-level features suitable for statistical analysis. Equipped with these facial embeddings, we proceed in three steps to predict credit outcomes, maintaining strict out-of-sample discipline.

First, we build ML models using only facial embeddings to predict borrower risk (Vantage scores and default probability), training solely on the extracted ArcFace embeddings, *excluding* all protected attributes. The resulting estimates represent predicted facial credit and facial delinquency scores. In parallel, we train a combined model that incorporates both Vantage scores and the facial embeddings as inputs to predict default probability, allowing the algorithm to capture nonlinear interactions between traditional and nontraditional

¹A main virtue of this construction is scale and granularity of information. The main limitation is external validity: because L2 covers registered voters in the U.S., our universe comprises individuals who were registered to vote at some point between 2017 and 2024; and because LinkedIn skews toward employed and connected workers, the sample leans slightly prime in credit quality. That said, along the two most policy-salient axes, sex and broad race/ethnicity bins in the voter file, the sample is broadly representative of national shares.

credit predictors. We benchmark the predictive ability of these two models against traditional credit-score-based predictions and against human capital measures from Revelio, such as salary and school rank.

Second, we use three-fold cross-validation to generate fully out-of-sample predictions for every borrower, ensuring no leakage between training and test sets.

Third, we examine heterogeneity in predictive performance across sex, income, age, and file thickness, linking these dimensions to settings where traditional scores are least informative and soft signals are plausibly most consequential for understanding credit risk.

Our main results are as follows. In the full sample, the ML-predicted facial credit score alone explains 12.7% of the variation in Vantage score. In contrast, salary and school rank – which are themselves novel labor market measures in the context of credit outcome predictions – explain 8.5% and 1.2% variation in credit scores, respectively. A joint model that includes the facial credit score plus age and demographic controls explains 19.1% of the variation in Vantage score, implying that demographics add incremental explanatory power, but that the facial information constitutes the majority of the joint model’s signal. Moving from the bottom to the top quintile of the facial-credit score distribution is associated with a 0.495 standard-deviation (46 points) higher predicted Vantage score, even after controlling for income, education, and demographics.

Consistent with this evidence, we find that facial embeddings also contain statistically and economically meaningful information about delinquency rates, both on their own and as an incremental signal. The facial delinquency score alone yields an AUC of 0.607, meaning the model correctly ranks a randomly selected defaulter above a randomly selected non-defaulter about 61% of the time. As above, the facial score has stronger predictive power than salary or education measures, in terms of standardized coefficients, AUC, and R^2 . Additionally, augmenting traditional credit scores (Vantage scores) with facial embeddings improves out-of-sample delinquency prediction by 1.5 percentage points in AUC.

Next, we document that the incremental predictive ability of the ML-based facial scores is concentrated precisely where traditional credit files are weakest. The gain in AUC for default prediction from the facial delinquency score is an order of magnitude larger – 6.4 percentage points – among thin-file borrowers (zero or one tradeline in 2017), versus only 0.8 percentage

points for thick-file borrowers. Similar patterns emerge across age and income: facial features add 5 to 8 percentage points of AUC for the youngest and lowest-income borrowers, declining monotonically to near-zero incremental value for older, wealthier individuals. These results suggest that facial embeddings proxy for residual information that is most salient precisely where traditional credit files provide the least signal.²

In a final set of analyses, we explore specific mechanisms that might underlie the predictive power of facial features as proxies for “soft” information. To do so, we train a separate ML model to predict *perceived* personality measures from facial features. Specifically, we use the ChaLearn First Impressions dataset which contains perceived Big 5 personality assessments from photos and videos rated by human evaluators, to train the model, and then apply it to the LinkedIn images in our sample. The Big 5 personality framework is the most widely used in psychology and economics and encompasses five broad traits: agreeableness, conscientiousness, extraversion, neuroticism, and openness.

We find that the perceived personality measures (the “Photo Perceived Big 5”) significantly predict both our facial credit measures and actual credit outcomes. For example, a one-standard-deviation increase in perceived agreeableness is associated with a 0.481 standard deviation increase in facial credit score and a 0.268 standard deviation increase in Vantage score. Perceived conscientiousness similarly predicts facial credit score and Vantage score positively, and both perceived agreeableness and conscientiousness predict lower facial delinquency scores and delinquency rates. These results suggest that perceived personality constitutes an important channel through which facial features encode credit-relevant “soft” information. More broadly, perceived personality may matter because it correlates with true latent personality (Connolly et al., 2007), shapes how others (including lenders) interact with borrowers, or reflects social support structures such as family or peer networks.

Taken together, our results provide three main insights. First, facial embeddings contain meaningful “soft” information about creditworthiness that is incremental to traditional credit

²We also examine differences by gender. Generally, the predictive effects are stronger for women than men: for example, facial features explain 20.3% of Vantage score variation for women versus 18.8% for men, with correspondingly larger “top 20% minus bottom 20%” dispersion (0.576 vs. 0.385 standard deviations). These larger effects for women emerge despite documented lower accuracy of ArcFace embeddings for female faces (Albiero et al., 2020, 2022, Bhatta et al., 2023), suggesting our estimates may be a lower bound on the true information content of facial appearance for female borrowers.

scores and human-capital proxies. Second, this incremental signal is strongest precisely where it is most valuable to the lenders (and borrowers): among thin-file, young, and lower-income borrowers for whom conventional scoring performs poorly. Third, perceived personality, as measured by perceived Big 5 traits extracted from photos, appears to be one mechanism through which facial features predict credit outcomes.

We reiterate that our goal is to evaluate the predictive power of facial embeddings in credit markets. Such tools could, in principle, reduce the regressive effects of traditional risk models, but we do not advocate the use of similar technologies in credit screening. While algorithms can be programmed to equalize score distributions across race, ethnicity, or gender, the harder ethical question is whether it is acceptable to rank individuals by facial features *within* demographic groups. Doing so risks violating autonomy and eroding respect for individuality, while weakening incentives to change behavior in ways not visible through facial traits. Ultimately, the ethical and welfare implications of using facial features in credit screening raise profound questions about the tension between technological capability and respect for human individuality. Nevertheless, in light of evolving industry practices and rapid advances in AI, independent academic evaluation is both timely and necessary.

The remainder of the paper proceeds as follows. Section 2 describes the data construction, including the Revelio-L2-Experian data, facial feature extraction, and Photo Perceived Big 5 measures. Section 3 outlines our estimation strategy, cross-validation procedures, and specifications. Section 4 presents the main results on credit score and delinquency prediction, heterogeneity analyses, and the perceived personality mechanism. Section 5 concludes.

2. DATA AND ML METHODOLOGY

We assemble an individual-level dataset by linking worker profiles (Revelio/LinkedIn), validated demographics from voter registration files (L2), and credit bureau files (Experian). The unit of observation is a person in the 2019 Experian vintage, augmented with that same person’s 2017 credit snapshot to form lagged variables. For each observation, we construct an anonymized key for their “full name-address-date of birth” to enable cross-source matching while preserving privacy. Each person has exactly one LinkedIn photo.

2.1 REVELIO LABS DATA

Our LinkedIn dataset comes from Revelio Labs, a leading provider of workforce data. These data include individuals’ education and employment histories, extracted from LinkedIn, along with profile images where available. Because the Revelio data lacks individuals’ addresses and dates of birth required for linkage to credit-bureau records, we match the Revelio/LinkedIn data to the L2 voter file. L2 is a continuously updated national voter-registration database compiled from state records covering more than 200 million registered voters. It provides a persistent person identifier and standardized demographic fields, including registration status, party affiliation, turnout, addresses/geocodes, contact information, and imputed demographics. Linking the LinkedIn data to L2 necessarily restricts our sample to individuals who have registered to vote in the United States.

Linking Revelio/LinkedIn to L2 proceeds in three stages. First, we standardize information in the LinkedIn profiles file by case-folding names, stripping punctuation, harmonizing addresses to USPS-style tokens when present, and recording dates of birth in YYYY-MM-DD format. If multiple snapshots exist for a person, we retain the most recent record as of 2019. We also inventory LinkedIn profile photos at this stage. Second, merge LinkedIn profiles to the L2 file to obtain validated demographics. To reduce false matches and improve computational efficiency, we perform record linkage within small “blocks” defined by on matches by last name, first name, age, and ZIP+4. Within each block, we retain the highest-confidence match based on fuzzy name similarity and exact match on year of birth. Ties are resolved by dropping the ambiguous pair rather than risking a false positive. The resulting intermediate Revelio-L2 merged file contains one person per row, with Revelio/LinkedIn fields complemented by L2 demographics (date of birth, permitted demographic variables, and location history). This intermediate merge yields 3,540,960 unique individuals. Third, we apply a large-language-model-based validation to the intermediate merged file, reclassifying 2,181,578 as true positives and the remainder as false matches. Of the 2,181,578 validated matches, 1,525,349 include a LinkedIn profile photo. Appendix A provides further details on all of data-cleaning for Revelio/Linked and L2, and merging steps between the two data sources.

2.2 EXPERIAN CREDIT DATA

Our initial sample consists of the 2,181,578 Revelio-L2 matches, of which 1,525,349 have a LinkedIn photo. We request Experian to match this 1,525,349 sample for our analysis. For each, we transmit standardized identifiers (first, middle, last, suffix, DOB, most recent address) and request annual snapshots. Experian successfully matched 1,495,792 individuals to internal records. We use two vintages: 2017 (for lagged credit variables) and 2019 (the analysis cross-section). Experian returns contemporaneous attributes (VantageScore, utilization rate, tradeline counts, and inquiries) and delinquency monitors summarizing performance over the preceding 6, 12, 24, and 36 months.

Variables-of-interest are defined as follows. Age, sex and ethnicity are provided in Revelio. Human-capital proxies come from Revelio: the log of raw salary, an indicator for missing school rank, and a numeric school-rank measure where higher values indicate worse prestige. Credit variables include the contemporaneous VantageScore, utilization, tradeline counts, and delinquency indicators constructed from Experian’s monitors. Our primary delinquency outcome equals one if the borrower experiences any 30-days-or-more past-due episode on any tradeline at some point in the twelve months immediately preceding the 2019 vintage date, conditional on being current twenty-four months prior (i.e., in 2017). As a robustness outcome, we instead code an indicator for whether the number of tradelines that are 30+ days past due increases over the twelve months preceding the 2019 vintage. These definitions exploit the bureau’s rolling monitors and avoid reconstructing event histories from raw tradeline panels. Finally, *Thin-file* status is coded from the 2017 snapshot: a borrower is thin if the number of trades in 2017 equals 0 or 1; otherwise thick.

2.3 ML FACIAL FEATURE EXTRACTION AND PREDICTIONS

We construct facial-feature-based predictors of credit outcomes in two steps: we first map each LinkedIn photo into a fixed-length facial embedding, and then feed these embeddings into a machine learning model to predict credit outcomes with strict out-of-sample evaluation.

We first extract facial features from LinkedIn profile photographs using an off-the-shelf,

state-of-the-art, deep convolutional neural network architecture specifically optimized for face recognition tasks. In this step, we convert high-dimensional, unstructured data – raw pixel values from photographs – into a structured, lower-dimensional feature representation that can serve as inputs to a predictive model. Our pipeline processes each image with a 50-layer residual network (ResNet-50) pretrained on millions of face images and fine-tuned using InsightFace’s ArcFace loss function (Deng et al., 2019, He et al., 2016). The network maps each face into a 512-dimensional embedding space in which Euclidean distances track perceptual similarity: faces of the same person cluster together, while faces of different people lie farther apart. For our setting with field photos, these embeddings are designed to be relatively invariant to lighting, head pose, expression, and image quality, while remaining sensitive to stable facial characteristics. We impose quality controls (for example, requiring exactly one detected face with sufficient detector confidence) to ensure stable detection of facial features for our large sample.

We construct facial-feature based predictions of credit outcomes to isolate incremental “soft” information inferrable from facial-features. Specifically, we predict three outcomes: the borrower’s contemporaneous Vantage score (continuous) and the two delinquency indicators defined earlier (binary). We consider two feature sets. In the *faces-only* specification, the model takes only the 512-dimensional embedding as input. In the *faces $\times$ Vantage* specification, we augment the embedding with the borrower’s 2017 Vantage score to allow for nonlinear interactions between traditional “hard” information and visually-inferred “soft” information.

To evaluate predictive performance, we partition the data into three folds. Each fold serves once as a held-out test set, while the union of the remaining two folds forms the training pool. For each individual in the dataset, we thus obtain a prediction from a model that was trained exclusively on other individuals. Within the training pool, we perform three-fold cross-validation, using an internal validation split and early stopping to tune hyperparameters.³ The test fold remains untouched until final scoring. This procedure

³We employ automated hyperparameter optimization with Optuna, which uses a Tree-structured Parzen Estimator (TPE) to search over a large hyperparameter space (tree depth, learning rate, number of boosting rounds, minimum samples per leaf, subsampling ratios over observations and features, and regularization parameters). The configuration that maximizes out-of-sample performance on the validation split is used for final training.

ensures that reported performance reflects genuine out-of-sample accuracy: the algorithm never sees test individuals at any stage of training or tuning.

Our primary modeling approach uses gradient-boosted decision trees implemented via LightGBM (Ke et al., 2017), a method widely recognized for strong performance on structured, high-dimensional data. For binary delinquency prediction, we focus primarily on the area under the receiver operating characteristic curve (ROC-AUC) as our main performance metric. We target precision-recall AUC to assess performance specifically at the high-risk (left) tail of the distribution, where classification decisions would typically concentrate in practice.

For continuous credit score prediction, we target three complementary metrics that are typical for prediction exercises: R^2 (the share of variance explained), root mean squared error (RMSE), and mean absolute error (MAE). The R^2 metric optimizes the fraction of the total variation in credit scores can be predicted from facial features. RMSE and MAE measure prediction accuracy in units of credit scores (Vantage points), with RMSE penalizing large errors more heavily than MAE. Together, these metrics offer a comprehensive assessment of how well facial features predict credit outcomes.

Together, these metrics provide a comprehensive view of how well facial features alone – and in combination with traditional scores – predict credit outcomes. Economically, the machine-learning model we train allows us to transform rich, unstructured images into a single statistic whose informational content in credit markets can be studied with standard methodologies.

2.4 PHOTO PERCEIVED BIG 5 PERSONALITIES

To obtain image-based measures of perceived personality, we train a deep learning model on 10,000 individuals from the ChaLearn First Impressions dataset, using a 6,000/2,000/2,000 train-validation-test split.⁴ Each ChaLearn video is first reduced to a single facial image. An automated face-selection module evaluates all frames on four quality criteria—sharpness (30%), illumination (30%), orientation (30%), and eye visibility (10%)—and retains the highest-scoring frame for each person. Extracted faces are aligned using 68 facial landmarks,

⁴See <https://chalearnlap.cvc.uab.cat/dataset/24/description/>.

resized to 224×224 pixels, and background-segmented so that the model focuses on the facial region rather than contextual cues.

We then map each face into a numerical representation using a standard feature-extracting deep neural network (InceptionResNetV1 pretrained on CASIA-WebFace). The network produces a 512-dimensional embedding that summarizes the visual features of the face. In the prediction stage, these embeddings are passed through a small feedforward network that outputs five normalized OCEAN scores – Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness. Appendix B provides additional implementation details and reports out-of-sample performance for our Photo Perceived Big 5 measures.

Intuitively, the model learns which patterns in the pixels co-move with human perceptions of these traits and distills that information into a low-dimensional score vector for each face. We then apply the trained model to our 1.1 million photos to obtain predicted Photo Perceived Big 5 personalities for all individuals in our main sample.

2.5 SUMMARY STATISTICS

The combined dataset for our analysis incorporates individual demographic information, human capital-proxies, rich-credit bureau data, feature-based predictions of credit outcomes, and Photo Perceived Big 5 measures for a panel of approximately 1.1 million consumers. Table 1 reports summary statistics for the person-level 2019 cross-section. We present the table for the full sample of individuals with a LinkedIn profile photo suitable for feature extraction and nonmissing credit-bureau data.

Relative to widely used bureau datasets, the sample skews young with median age of 28, and relatively high-income with a median salary of \$66k (Lee and van der Klaauw, 2010, Gibbs et al., 2025). Both the average and median credit quality are prime (mean VantageScore=713; P50=729; P90=817); only 7% have a 30+ days-past-due (DPD) delinquency over the past 12 months as of 2019, conditional on being current in 2017. Using commonly referenced VantageScore bands (prime $\approx$ 661-780), our borrowers are generally in the prime range (Goodman, 2024). Eighteen percent of the sample meets our thin-file definition (0-1 trades in 2017). This is not surprising given that our sample includes individual with a populated LinkedIn profile. For context, prior work often uses ≤ 3 accounts when dis-

cussing noise and information sparsity in credit reports (Blattner and Nelson, 2021). These composition facts matter for interpretation: falling below mortgage lenders’ minimum score thresholds has large, persistent effects on borrowing (Laufer and Paciorek, 2022), so a sample concentrated above typical thresholds will exhibit low realized delinquency conditional on approval.

Finally, our “facial credit score” should be viewed as incremental to bureau data, consistent with evidence that nontraditional signals and digital footprints complement—and sometimes rival—bureau scores in predicting default or shaping credit outcomes (Duarte et al., 2012, Ravina, 2020, Berg et al., 2020, Bartlett et al., 2022, Fuster et al., 2022).

Figure 1 establishes that traditional credit scores correlate strongly with human capital measures. Panel (a) shows a positive relationship between Vantage scores and log salary, with a slope of 48.9 points per log unit. Panel (b) demonstrates that better school rank (lower numerical values) associates with higher Vantage scores, with a slope of -0.052. These patterns confirm that credit scores capture economically meaningful variation in borrower quality, providing a baseline against which to evaluate facial feature-based predictions.

3. EMPIRICAL APPROACH

3.1 FACIAL FEATURES AND CREDIT OUTCOMES

We first test whether standardized high-dimensional features extracted from faces predict credit-related outcomes out-of-sample, and then ask what that predictive content means economically once standard proxies for human-capital are held constant.

We benchmark the performance of our feature-based outcomes in predicting out-of-sample delinquency behavior against lagged traditional credit scores. The former is our measure for “soft” information and the latter is a widely-used measure for “hard” information. Model comparisons emphasize differences in AUC and changes in R^2 . These metrics are descriptive: they document whether facial embeddings contain structured signal about credit quality, and how that signal interacts with the information already embedded in credit-bureau scores.

To translate algorithmic prediction into economics, we estimate cross-sectional regressions in the 2019 sample. Let p_i^{faces} denote the model-implied predictions for Vantage scores in 2019

or future delinquencies. We regress the realized outcomes on p_i^{faces} and then add the lagged Vantage score and the human-capital controls described above: log salary, an indicator for missing school rank, the numeric school rank. All specifications use heteroskedasticity-robust standard errors.

We estimate the following,

$$y_i = \alpha + \alpha_{j(i)} + \beta \cdot p_i^{faces} + \mathbf{\Gamma}' \mathbf{X}_i + \varepsilon_i \quad (1)$$

where y_i are credit outcomes of interest (e.g., Vantage score in 2019, new delinquencies in 2019 for borrowers that were current in 2017, etc.), $\alpha_{j(i)}$ are fixed effects for sex and ethnicity, and X_i is a vector of additional controls, including linear and quadratic age and a set of benchmarking covariates (e.g., education, income, lagged Vantage score). All variables with a (Z) are standardized. The coefficient of interest is β , which captures the predicted change in credit outcomes for a one standard deviation change in our facial features-based prediction. We compare this coefficients to those of other established predictors of credit market outcomes, such as income and education. These comparisons allow us to conclude, for example, that a facial features-based prediction possess predictive power comparable to commonly employed benchmarks in industry and academia.

We report R^2 for all regression specifications as a compact summary of in-sample explanatory power. For binary delinquency outcomes, we complement R^2 with a standard out-of-sample (OOS) classification performance test: the receiver operating characteristic and its area under the curve (AUC). By construction, AUC ranges from 0.50 (random classification) to 1.00 (perfect classification), and admits a probability interpretation: it is the chance that the model ranks a randomly drawn delinquent borrower as riskier than a randomly drawn non-delinquent borrower (Hanley and McNeil, 1982). Following the credit-risk literature, we use AUC as our headline measure of efficacy of various models that rely on tradition versus non-traditional predictors (Altman et al., 2010, Iyer et al., 2016, Vallée and Zeng, 2019). As a calibration, the Vantage-only benchmark in our data attains an AUC of 0.677, which is astonishingly similar to the AUC of 0.683 achieved by the traditional credit scoring model in consumer credit in Germany by (Berg et al., 2020). In information-sparse

settings, AUCs around 0.60 are often viewed as practically useful, while values ≥ 0.70 are typical in information-rich environments (Iyer et al., 2016).

We also estimate equation (1) separately by sex. The goal is to assess whether the incremental predictive content of facial embeddings is similar for men and women, not to propose group-specific models. A large literature documents sex differences in credit market outcomes even for otherwise-comparable borrowers across several products such as small-business loans, overdrafts, mortgages and consumer platforms (Pope and Sydnor, 2011, Fang and Munneke, 2020, Alesina et al., 2013, Brock and De Haas, 2023, Goldsmith-Pinkham and Shue, 2023). Even though sex never enters our prediction models, differences in base rates or outcome variance can make pooled coefficients difficult to interpret. Estimating within-sex provides a simple diagnostic: it shows whether the slope on p_i^{faces} is stable across groups or whether the same facial signal translates into systematically different credit outcomes. We view this as a robustness and measurement check rather than a prescriptive modeling choice.

Finally, we compute AUCs not only for the full sample but also within narrow bins of variables that proxy for data richness: income, age, and file thickness. For each bin, we compare two models: a traditional benchmark using Vantage alone and a model that augments Vantage with facial embeddings. This design allows us to ask a simple question of economic interest: are traditional credit scores most informative for wealthier, older, thick-file borrowers, and least informative for lower-income, younger, thin-file borrowers? If so, we should see the incremental AUC from adding facial embeddings shrink in information-rich cells and expand in information-sparse cells. In this sense, subgroup AUCs serve to characterize how regressive traditional scores are in practice and where image-based soft information has the highest marginal value.

3.2 FACIAL FEATURES AND PERCEIVED PERSONALITY TRAITS

To better understand the mechanisms through which facial features encode credit-relevant information, we examine whether our facial measures capture behavioral traits associated with creditworthiness. The psychological literature establishes that personality traits – particularly those captured by the Big 5 taxonomy – significantly predict financial behaviors including savings patterns, debt accumulation, and default propensity (Donnelly et al., 2012,

Brown and Taylor, 2014). Conscientiousness correlates positively with creditworthiness through its association with planning, self-discipline, and rule-following behavior. Agreeableness relates to cooperation and contract adherence. Neuroticism captures emotional stability and stress management, factors relevant to financial decision-making under economic uncertainty.

Recent advances in computer-vision demonstrate that neural networks can extract personality impressions from facial images with substantial predictive validity (Celiktutan et al., 2014). These “first impression” scores reflect rapid personality judgments that humans make based on facial appearance. While such judgments need not be accurate measures of true personality, they may still influence social and economic interactions, including lending relationships. If lenders or other market participants form systematic impressions of borrower reliability based on appearance, and if borrowers’ behavior partly reflects or responds to these impressions, then facial features may encode personality-related information relevant for credit outcomes.

We discipline this mechanism test in a simple way. Specifically, we investigate whether standardized Photo Perceived Big 5 personality scores explain variation in both (i) actual credit outcomes (Vantage scores and delinquency) and (ii) our facial credit measures. This dual test allows us to distinguish whether our facial embeddings capture personality-relevant information that is itself predictive of creditworthiness, or instead reflect orthogonal variation in facial appearance unrelated to behavioral traits.

We construct Photo Perceived Big 5 scores using state-of-the-art computer-vision models trained to predict personality impressions from faces. We described our protocol in detail earlier. We estimate regressions that relate these scores to both actual credit outcomes and our facial measures. For credit scores, we estimate:

$$y_i = \alpha + \beta_1 \cdot \text{Agreeableness}_i + \beta_2 \cdot \text{Conscientiousness}_i + \beta_3 \cdot \text{Extraversion}_i + \beta_4 \cdot \text{Neuroticism}_i + \beta_5 \cdot \text{Openness}_i + \mathbf{\Gamma}' \mathbf{X}_i + \varepsilon_i \quad (2)$$

where y_i is either a realized or facial-feature-based credit outcome (Vantage score or delinquency), all personality variables are standardized to mean zero and unit variance, X_i de-

notes a vector of demographic controls (sex $\times$ ethnicity fixed effects, age, and age squared). The dependent variables in the delinquency specifications are binary indicators for payment delinquency. This specification isolates the incremental explanatory power of perceived personality traits beyond standard demographic characteristics.

The test provides two insights. First, if Photo Perceived Big 5 scores predict actual credit outcomes, this indicates that appearance-based personality impressions contain information about borrower behavior. Second, if these scores also predict our facial credit measures, this suggests that our machine-learning models extract personality-relevant features from facial embeddings – a channel through which faces may encode signals of creditworthiness.

4. RESULTS

4.1 FACIAL FEATURES PREDICTING VANTAGE SCORES

We begin by examining whether facial embeddings contain economically meaningful information about borrowers’ credit quality. As a first step, we relate our facial-features-based credit score to Vantage scores, both unconditionally and after controlling for demographic and human-capital controls. This analysis provides a baseline assessment of the explanatory power of photo-based credit quality and benchmarks the contribution relative to traditional correlates of creditworthiness.

Table 2 reports the relationship between facial features-based credit scores and Vantage scores. Panel A shows in columns (1)–(3) that demographics (sex, ethnicity, and age) jointly explain just over 16% of the variation in Vantage scores. In column (4), the facial credit score *alone* explains 12.7% of the variation. Including demographics and age controls in column (5) increases the R^2 to 19.1%, suggesting that the facial credit score captures substantial, economically meaningful variation in creditworthiness beyond basic demographic characteristics. The standardized coefficient on the facial credit score is 0.356 without controls and 0.199 with full controls corroborating the economic relevance of the measure. Moving from the bottom to the top quintile of the facial credit score distribution is associated with a 0.99 standard-deviation difference in Vantage scores without controls, and a 0.55 standard-deviation difference with full controls, a sizable change in credit quality.

Panel B adds human-capital controls to assess the extent to which facial features proxy for traditional correlates of creditworthiness such as income and education. column (1) shows that the baseline specification with just demographics and age achieves an R^2 of 14.9%. Column (2) shows that facial features explain 11.8% of variation in credit scores. Columns (3) and (4) show that log salary alone explains 8.5% of variation, while school rank explains 1.2%. In column (6) combining facial credit score with salary, school rank, and full controls increases the R^2 of 19.7%, with the facial credit score maintaining a standardized coefficient of 0.181. This indicates that facial features provide incremental predictive power beyond standard economic proxies. In this specification, individuals in the top quintile of the facial credit score distribution have Vantage scores that are 0.495 standard deviations higher than those in the bottom quintile.

Table 3 examines whether these patterns differ by sex. Panel A presents results for women. The results mirror those for the overall sample in Table 2, but suggest a slightly stronger relationship. Specifically, the facial credit score has a standardized coefficient of 0.210 with full controls (column (6)), which is slightly higher than in the pooled sample, indicating that the association between facial features and creditworthiness is more pronounced for female borrowers. The model achieves an R^2 of 20.3% for women, and moving from the bottom to the top quintile of the facial credit score distribution corresponds to a 0.576 standard-deviation gap in Vantage scores, notably larger than in the full sample.

Panel B shows the results for men. The facial credit score coefficient is 0.140 with full controls (column (6)), which is smaller than for women. The R^2 reaches 18.8%, and the the corresponding top-bottom quintile gap is 0.385 standard deviations. These results indicate that facial features explain more variation in credit scores for women than for men. This pattern arises despite prior evidence that ArcFace embeddings – the facial recognition technology we employ – exhibit systematically lower accuracy for women than men (Albiero et al., 2020).⁵ That our estimates for women are stronger despite likely higher measurement error in facial features for female borrowers, suggesting the true relationship between facial

⁵Recent computer-vision research has established that ArcFace and similar deep learning face recognition systems exhibit lower accuracy for women, with both higher false match rates and higher false non-match rates driven by factors including hairstyle occlusion, makeup variation, and facial morphology differences (Albiero et al., 2022, Bhatta et al., 2023).

appearance and creditworthiness for women may be even larger than what we estimate.

Overall, these results show that facial images contain meaningful information about borrowers' credit scores that is incremental to standard demographic and human-capital controls.

4.2 FACIAL FEATURES PREDICTING DELINQUENCY

We now turn from score-level prediction to examining whether the image-derived signals that track credit scores also help to identify borrowers who become delinquent. We evaluate out-of-sample classification performance (ROC-AUC and related metrics), test incremental value over lagged Vantage scores and human-capital controls, and examine heterogeneity by file thickness, age, income, and sex. We define a delinquency as 30-days-or-more past-due episode on a tradeline during the 12 months immediately preceding t (the window $[t - 12, t]$), conditional on being current at $t - 24$.⁶

Table 4 Panel A reports the classification results. Column (3) provides the benchmark specification with demographic fixed effects and age controls, which attains an AUC of 0.591 and an R^2 of 0.7%. Column (1) shows that the facial delinquency score alone attains a higher AUC of 0.607 and explains 1.1% of the variation in delinquency. Combining the facial score with demographic and age controls in column (5) increases the AUC to 0.626 and the R^2 to 1.4%, indicating that the facial score provides incremental explanatory power for delinquencies beyond standard demographic characteristics. The top–bottom quintile gap in this specification is 6.1 percentage points, reflecting meaningful dispersion in predicted delinquency risk across image-derived scores.

Panel B incorporates human-capital controls. Columns (1)–(2) reproduce the key specifications from Panel A. Columns (3)–(4) show that log salary alone yields an AUC of 0.520, while school rank yields an AUC of 0.544. In column (5) we combine the facial delinquency score with salary and school rank and attain an AUC of 0.620, representing an improvement of roughly 7–8 percentage points over the human-capital-only models. In column (6), com-

⁶All delinquency results that follow in this and subsequent sections are equally robust to an alternate definition of delinquency and are presented in Appendix C. In those results, *Delinquency* is defined as an indicator equal to 1 if the borrower experiences an increase in the number of tradelines that are 30-days-or-more past-due in the 12 months immediately preceding t (the window $[t - 12, t]$).

binning the facial delinquency score with salary, school rank, and the full set of controls yields an AUC of 0.629 and an R^2 of 1.5%. In this full specification, the standardized coefficient on the facial delinquency score is 0.021, and top–bottom quintile gap is 5.8 percentage points, indicating economically meaningful dispersion in predicted delinquency risk.

In Table 5 we add lagged Vantage scores (from 2017) to assess whether facial features provide information beyond traditional credit scores. Panel A shows that adding lagged Vantage score (column (3)) substantially improves performance, achieving AUC of 0.677 and R^2 of 2.3%. Augmenting this specification with the facial delinquency score in column (5) increases the AUC to 0.722 and R^2 to 4.4%, which provides evidence of meaningful incremental predictive content in facial features beyond what traditional scores capture. In this specification with full controls, the standardized coefficient on the facial delinquency score is 0.014.

Panel B adds human capital controls to the specification. Column 7 shows that the full model (facial delinquency, salary, school rank, and lagged Vantage) achieves AUC of 0.724 and R^2 of 4.5%. The standardized coefficient on the facial delinquency score in the specification with full controls (Column 7) is 0.014. As before, we find meaningful predictive content in facial features-based predictions beyond traditional scores and human capital proxies.

Table 6 presents delinquency predictions separately by sex. Panel A (women only) shows that facial delinquency score combined with lagged Vantage and full controls (Column 7) achieves AUC of 0.720 and R^2 of 4.7%. The facial delinquency coefficient is 0.015. The *Top20-Bottom20* gap is 4.2 percentage points. Panel B (men only) shows that the same specification (Column 7) achieves AUC of 0.726 and R^2 of 4.3%. The facial delinquency coefficient is 0.011, notably smaller than for women. The *Top20-Bottom20* gap is 0.029 standard deviations.

Consistent with the Vantage score results, facial features explain larger variation in outcomes for women (coefficient of 0.015 vs 0.011, *Top20-Bottom20* of 0.042 vs 0.029), suggesting that facial information is particularly valuable for predicting female borrower behavior. Once again, the fact that facial features demonstrate stronger predictive relationships with credit outcomes for women – despite documented measurement noise in recognition systems

– suggests our results represent a lower bound on the true information content of facial appearance for female borrowers.

4.2.1 PREDICTIVE PERFORMANCE BY CREDIT FILE THICKNESS

We compare the predictive performance of traditional credit scores (“hard” information) and facial-feature-based (“soft” information) across thin-file and thick-file borrowers.

Table 7 reports results for the full sample split by file thickness. Panel A focuses on the facial delinquency score *alone*. For thick-file borrowers (Columns 1–2), the facial delinquency score with full controls attains an AUC of 0.610. The corresponding coefficient on the facial delinquency score is 0.025, and the base delinquency rate is 7.6%. For thin-file borrowers (Columns 3–4), the AUC is higher at 0.668, with a coefficient of 0.015, and a base delinquency rate of 3.8%.

Panel B adds lagged Vantage scores and allows a direct comparison with “hard” information. In the specification with lagged Vantage but without facial features, the AUC is 0.741 for thick-file borrowers and 0.673 for thin-file borrowers (Columns 2 and 5), with R^2 of 5.8% and 1.5%, respectively. Adding the facial delinquency score for thick-file borrowers (Columns 2 vs. 3) raises the AUC from 0.741 to 0.744 (0.3 percentage points) and R^2 from 5.8% to 6.0%. For thin-file borrowers (Columns 5 vs. 6), the AUC increases from 0.673 to 0.695 (2.2 percentage points) and R^2 from 1.5% to 2.0%. In this full thin-file specification, the coefficient on the facial delinquency score is 0.013.

These results demonstrate that facial features provide much larger incremental information than traditional credit score for borrowers with sparse credit histories.

Table 8 explores file thickness heterogeneity separately by sex. Panel A (women only) shows that for thick-filed women, adding facial delinquency to lagged Vantage improves AUC from 0.735 to 0.739 (0.4 percentage points), with R^2 rising from 5.9% to 6.1%. For thin-filed women, the AUC improvement is from 0.673 to 0.698 (2.5 percentage points), with R^2 rising from 1.6% to 2.2%. The facial delinquency coefficient is 0.015 for thin-filed women compared to 0.014 for thick-filed women. The *Top20-Bottom20* dispersion is notably larger for thin-filed women (0.042 vs. 0.039).

Panel B (men only) shows that for thick-filed men, the AUC improvement from adding

facial features is from 0.747 to 0.749 (0.2 percentage points), with R^2 rising from 5.5% to 5.7%. For thin-filed men, the improvement is from 0.667 to 0.688 (2.1 percentage points), with R^2 rising from 1.3% to 1.7%. The facial delinquency coefficient is 0.011 for thin-filed men versus 0.010 for thick-filed men. The *Top20-Bottom20* dispersion is larger for thin-filed men (0.030 vs. 0.028).

Across both sexes, facial features provide substantially more incremental value for thin-filed borrowers, though the absolute magnitudes of coefficients and dispersions remain larger for women than men within each file thickness category.

4.2.2 PREDICTIVE PERFORMANCE ACROSS AGE, INCOME, AND CREDIT HISTORY

We next examine whether the pattern documented by file thickness – “hard” information being strongest where credit files are rich and facial features adding more where they are sparse – extends across other characteristics of borrowers that are typically associated with credit thickness: number of age, income, and credit history (number of tradelines).

Figures 2, 3, and 4 present AUC performance across borrower characteristics as of $t - 2$. Each figure plots three series: facial features only (red circles), Vantage scores only (green squares), and facial features combined with Vantage scores (yellow diamonds), all estimated with demographic and age controls.

Figure 2 shows performance by number of tradelines (credit history thickness). At 0 tradelines, Vantage scores alone attain an AUC of about 0.625, while the facial-features-only and combined models both attain AUCs of about 0.675 – a gap of roughly 5 percentage points relative to Vantage scores alone. As the number of tradelines increases, Vantage scores’ performance rises substantially (reaching about 0.76 at 6 tradelines), while the facial-features-only model remains around 0.625. The combined model remains slightly above Vantage scores throughout (around 0.755–0.760). For borrowers with few tradelines, facial features outperform traditional credit scores, and combining the two yields sizable gains; for borrowers with thicker files (4+ tradelines), Vantage scores dominate the facial-features-only model and the incremental gain from combining is small.

Figure 3 displays performance by borrower age. At age 18, Vantage scores attain an AUC of about 0.66, while the facial-features-only and combined models attain AUCs near 0.74 (a

gap of roughly 8–9 percentage points relative to Vantage). As age increases, performance converges: by age 38, the facial-features-only model attains an AUC of about 0.60, while Vantage scores and the combined model both attain AUCs of about 0.74. The AUC difference between Vantage scores and the combined model is largest for the youngest borrowers (around 8–9 percentage points) and near zero for older borrowers. This indicates that facial features provide the most incremental information for young borrowers who have not yet accumulated extensive credit histories.

Figure 4 shows performance by borrower income. At a salary of \$25,000, facial features achieve AUC of 0.66, Vantage scores achieves 0.69, and the combined model achieves 0.75. At \$150,000 salary, the values converge: facial features achieve 0.58, Vantage scores and the combined model achieve 0.72. The incremental gain from facial features declines monotonically from around 6 percentage points for the lowest income group to under 1 percentage point for the highest income group. This pattern suggests facial features are most informative for lower-income borrowers whose traditional credit files may be less predictive of behavior.

Figures 5, 6, and 7 repeat the analysis separately for women (Panel a) and men (Panel b). Figure 5 shows tradeline heterogeneity by sex. For women with 0 tradelines, the combined model outperforms the traditional credit scores by 5-6 percentage points. For men with 0 tradelines, this gain is close to 2-3 percentage points. Figure 6 displays a similar pattern across the age distribution by sex, and Figure 7 shows a similar pattern across the income distribution by sex. Women’s Vantage-only scores systematically underperform men’s throughout the age and income distribution.

Figure 8 presents a complementary analysis using school rank instead of facial features. Across the income distribution, the combined model using school rank plus Vantage almost perfectly overlaps with the Vantage-only curve, indicating minimal incremental value from school rank at any income level. In contrast with the facial-features case, there is no comparable income-related pattern in predictive gains when using education as an additional predictor.

4.2.3 ROC COMPARISONS BY CREDIT FILE THICKNESS

Studying ROC curves complement our AUC analysis by credit file thickness. We show the full trade-off between true and false positive rates at all possible thresholds, which allows us to see where along the risk spectrum the combined model improves on Vantage scores and how these gains differ between thin- and thick-file borrowers.

Figure 9 presents ROC curves comparing the Vantage-only model (blue line) to the Features+Vantage combined model (red line) for the full sample. The Vantage-only model achieves an AUC of 0.674, while the combined model achieves an AUC of 0.692, an improvement of 1.8 percentage points. The χ^2 test statistic of 809.51 confirms that this difference is highly statistically significant ($p < 0.001$) based on 882,442 observations.

Figure 10 splits the ROC comparison by file thickness. For thin-file borrowers (left panel), Vantage alone achieves an AUC of 0.607 while Features+Vantage achieves 0.671, a gain of 6.4 percentage points ($\chi^2 = 424.12$, $p < 0.001$, $N = 165,003$). For thick-file borrowers (right panel), Vantage alone achieves an AUC of 0.727 while Features+Vantage achieves 0.735, a gain of 0.8 percentage points ($\chi^2 = 455.37$, $p < 0.001$, $N = 717,439$). This eight-fold difference in AUC improvement (6.4 vs. 0.8 percentage points) is visible in the ROC curves: the combined model dominates Vantage much more strongly for thin-file borrowers than for thick-file borrowers.

Table 9 formally tests whether the AUC differences between thin and thick files are statistically significant. The table reports differences in true positive rate (TPR) at fixed false positive rates (FPR) for the full sample, thin files, and thick files. At FPR of 0.10, the Features+Vantage model achieves a 3.1 percentage point higher TPR than Vantage alone in the full sample, indicating that the combined model outperforms the Vantage alone model in regions of low FPR. For thin-file borrowers, this advantage is 12.2 percentage points, while for thick-file borrowers it is -1.4 percentage points, indicating Vantage slightly outperforms the combined model at this threshold. At FPR of 0.50, the thin-file advantage is 5.9 percentage points versus 0.0 points for thick files. At FPR of 0.90, thin files show a -0.7 percentage point difference (Vantage alone outperforms) while thick files show a $+0.5$ percentage point difference (the combined model outperforms).

These results confirm that the incremental value of facial features is concentrated at lower levels of FPR and for thin-filed borrowers. The advantage wears off as we progress towards high levels of FPR or towards tick-filed borrowers. Lower FPR regions are typically when the approval threshold for lenders is relatively low, implying adding facial features is more beneficial when a larger pool of applications is expected to be approved than denied.

In a further test of robustness of our facial-features-based prediction results, we replicate our analysis for a predominant borrower group within a sex-gender pool. Doing so, our prediction model is fully purged of any demographic differences across borrower groups that could be correlated with facial features. We report our findings in Appendix D. We find that our analysis within this borrower group yields findings that are conclusively similar to our full sample of borrowers.

In summary, facial feature-based predictions provide incremental information beyond traditional credit scores, with effect sizes comparable to major economic fundamentals such as income and education. This incremental value is concentrated among borrowers with thin credit files, younger individuals, lower-income earners, and women. For these groups, augmenting traditional credit scores with facial features improves delinquency prediction by 2–6 percentage points of AUC. For borrowers with thick credit files, higher incomes, and older ages – precisely those for whom traditional scoring already performs well – facial features add little additional predictive power. The incremental value of facial features is thus largest where conventional “hard” information is scarcest.

4.3 PHOTO PERCEIVED PERSONALITY AND CREDIT OUTCOMES

We use Photo Perceived Big Five scores to test a natural mechanism for why faces predict credit outcomes: if facial embeddings partly compress appearance-based personality impressions, those impressions should predict both actual credit outcomes and our facial-based credit measures.

Table 10 regresses credit scores and facial credit measures on Photo Perceived Big Five personality traits. Panel A examines Vantage scores; Panel B examines our constructed facial credit scores.

Examining actual credit outcomes first, we find that four of the five Big5 traits exhibit

highly significant associations with Vantage scores (Column 1). Agreeableness shows the strongest effect, with a one-standard-deviation increase associated with a 0.268 standard deviation increase in Vantage score ($p < 0.01$). Conscientiousness (0.178σ , $p < 0.01$), Extraversion (0.116σ , $p < 0.01$), and Openness (0.203σ , $p < 0.01$) all predict higher credit scores. Neuroticism exhibits a negative association (-0.652σ , $p < 0.01$), consistent with the hypothesis that emotional instability correlates with financial distress and payment difficulties. These effects remain after controlling for sex $\times$ ethnicity fixed effects and a flexible specification in age, suggesting that perceived personality traits capture variation in creditworthiness beyond observable demographics. The model achieves an R^2 of 0.169, indicating that appearance-based personality impressions combined with demographic controls explain approximately 17% of the variation in credit scores.

Panel B (Column 2) shows that Photo Perceived Big Five scores also predict our facial credit scores, with similar sign patterns and larger magnitudes. Agreeableness (0.481 , $p < 0.01$), Conscientiousness (0.226 , $p < 0.01$), and Openness (0.199 , $p < 0.01$) are positively associated, while Neuroticism is strongly negatively associated (-0.765 , $p < 0.01$). Extraversion is near zero and not statistically significant (0.003 , $p > 0.10$). The R^2 of 0.322 indicates that perceived personality explains about one-third of the variation in facial credit scores in the presence of demographic controls. Together, Panel A and Panel B show that the same appearance-based personality impressions that predict actual credit scores also predict what our facial model learns.

Table 11 extends this analysis to delinquency outcomes, providing converging evidence on the personality-behavior link in credit markets. Panel A examines whether Photo Perceived Big 5 scores predict actual 90-day delinquency episodes in the 12 months following measurement. All five personality traits exhibit statistically significant associations in the theoretically expected directions. More agreeable individuals show lower delinquency propensity (-0.036σ , $p < 0.01$), as do more conscientious (-0.013σ , $p < 0.01$), extraverted (-0.008σ , $p < 0.01$), and open (-0.018σ , $p < 0.01$) borrowers. Conversely, Neuroticism positively predicts delinquency (0.066σ , $p < 0.01$), consistent with the interpretation that perceived emotional instability correlates with elevated default risk. While the R^2 of 0.008 appears modest, this is unsurprising given the challenging nature of predicting binary rare events

(delinquency rates in our sample average approximately 10 – 15%) from visual information alone. The AUC of 0.596 indicates that personality impressions from photographs achieve decent classification power.

Most importantly, Panel B demonstrates that these same personality traits significantly predict our facial delinquency measure. The coefficient patterns closely mirror those for actual delinquency, with Agreeableness (-0.011σ , $p < 0.01$), Conscientiousness (-0.003σ , $p < 0.01$), Extraversion (-0.001σ , $p < 0.05$), and Openness (-0.006σ , $p < 0.01$) all negatively predicting facial delinquency scores, while Neuroticism exhibits a positive association (0.017σ , $p < 0.01$). R^2 of 0.110 suggests that our facial embeddings successfully encode personality-relevant variation that relates to delinquency propensity, but there remains other unobserved explanatory power in the facial measure beyond our perceived personality measures.

Table 12 repeats the delinquency analysis separately for thin- and thick-file borrowers. The sign and relative magnitudes of the Big 5 coefficients are similar to the full-sample results. Consistent with earlier findings, adding nontraditional information increases AUC by about 6 percentage points for thin-file borrowers, with smaller gains for thick-file borrowers.

These results carry several important implications for interpreting our main findings. First, the consistent patterns across actual outcomes and facial measures – with nearly identical sign structures and ordering of coefficient magnitudes – provide evidence that our machine learning models extract meaningful personality-related information from facial features. The facial credit and delinquency scores are not merely capturing arbitrary visual variation; rather, they appear to encode appearance-based signals systematically related to behavioral traits known to influence financial decision-making.

Second, the magnitude of effects suggests that personality impressions represent a meaningful component of the information contained in facial features. The fact that Photo Perceived Big 5 scores explain 32% of the variation in facial credit scores (compared to 17% for actual Vantage scores) indicates that appearance-based personality judgments may constitute a primary channel through which facial features encode credit-relevant information. This finding aligns with psychological theories of “thin slice” judgments, wherein individuals form rapid impressions of others’ traits from minimal visual information – impressions

that, while imperfect, contain statistical validity (Ambady and Rosenthal, 1992, Ludwig and Mullainathan, 2024).

Third, the results speak to potential mechanisms underlying the predictive power of facial features in credit markets. One interpretation holds that facial appearance genuinely correlates with personality traits, which in turn influence financial behaviors. Research in evolutionary psychology and behavioral genetics documents modest but significant associations between facial morphology and personality (Penton-Voak et al., 2006). An alternative interpretation emphasizes self-fulfilling prophecies: if individuals are consistently treated in accordance with personality impressions formed from their appearance, they may gradually conform to such expectations through a process of behavioral confirmation (Snyder, 1984). Our data cannot definitively adjudicate between these mechanisms, but both pathways would produce the observed patterns wherein Photo Perceived Big 5 scores predict both actual behavior and facial measures.

Fourth, the differential predictive power across personality dimensions provides insight into what facial features encode. Neuroticism exhibits the strongest associations with both credit scores (in absolute magnitude) and delinquency outcomes. This may reflect the fact that certain facial characteristics – perhaps related to perceived trustworthiness, stress, social networks, or emotional control – provide particularly salient signals about an individual’s financial reliability. The weaker or null results for Extraversion in some specifications suggest that not all personality dimensions are equally encoded in or extractable from facial appearance using current computer vision methods.

We emphasize three important caveats in interpreting our findings. First, Photo Perceived Big 5 scores measure perceived rather than actual personality. The models predict how observers would rate individuals’ personalities based on appearance, not individuals’ true psychological traits as measured through validated self-report or behavioral assessments. The predictive power of these scores for credit outcomes could reflect either: (a) accurate inference of personality from faces, or (b) social processes wherein appearance-based expectations influence opportunities, interactions, and ultimately behaviors. Disentangling these channels would require matched data on actual personality measures.

Second, while the Photo Perceived Big 5 model achieves state-of-the-art performance on

benchmark datasets ($R^2 = 0.326$, $\text{MAE} = 0.096$ on ChaLearn First Impressions), substantial measurement error likely remains in the personality impression scores. This measurement error would attenuate our estimated coefficients, suggesting that the true relationships between personality impressions and credit outcomes may be stronger than our point estimates indicate.

Third, the modest R^2 values in the delinquency regressions remind us that personality impressions, while statistically significant, explain only a small fraction of default risk and our facial measures. Many factors such as health emergencies, stress from job or family situations – that are unrelated to stable personality traits – but can affect facial appearance, and add informativeness towards actual credit outcome. The Photo Perceived Big 5 scores provide incremental information.

Despite these limitations, the evidence that standardized, algorithmically-generated personality impressions from photographs systematically predict both credit outcomes and our facial measures strengthens confidence in our main findings. The results suggest that facial features encode meaningful behavioral information – information that can be captured through machine learning and that relates to traits known to influence financial decision-making. This mechanistic evidence complements our baseline results and provides a foundation for future research examining the psychological and social processes through which appearance influences economic outcomes.

5. CONCLUSION

Using data on more than one million U.S. borrowers, this paper shows that standardized facial embeddings extracted from LinkedIn profile photographs contain incremental, out-of-sample predictive signal about consumer creditworthiness, with this information concentrated precisely where traditional credit scores are least informative. Augmenting traditional credit scores with facial features improves delinquency prediction by 6.4 percentage points of AUC for thin-file borrowers, compared to less than one percentage point for thick-file borrowers, with similar patterns across age and income gradients. We provide evidence that one mechanism through which this predictive ability operates is behavioral tendencies, as

measured by perceived personality traits extracted from photos as well.

These findings suggest facial information could help reduce measurement error in settings where “hard” information is sparse, potentially expanding credit access for underserved populations. However, the use of facial data in lending raises profound questions about privacy, consent, algorithmic transparency, and disparate impact that are beyond the scope of this paper. Instead, we view our results as documenting an empirical regularity – that faces contain substantial “soft” information about creditworthiness – rather than advocating for any particular policy. Whether and how such information should be used remains an open question for regulators, lenders, and society.

REFERENCES

- Albiero, V., K. S. Krishnapriya, K. Vangara, K. Zhang, M. C. King, and K. W. Bowyer (2020). Analysis of gender inequality in face recognition accuracy. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops*, pp. 81–89.
- Albiero, V., K. Zhang, M. C. King, and K. W. Bowyer (2022). Gendered differences in face recognition accuracy explained by hairstyles, makeup, and facial morphology. *IEEE Transactions on Information Forensics and Security* 17, 127–137.
- Alesina, A. F., F. Lotti, and P. E. Mistrulli (2013). Do women pay more for credit? evidence from italy. *Journal of the European Economic Association* 11(S1), 45–66.
- Altman, E. I., G. Sabato, and N. Wilson (2010). The value of non-financial information in SME risk management. *Journal of Credit Risk* 6(2), 95–127.
- Ambady, N. and R. Rosenthal (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin* 111(2), 256–274.
- Athey, S., D. Karlan, E. Palikot, and Y. Yuan (2022, Nov). Smiles in profiles: Improving fairness and efficiency using estimates of user preferences in online marketplaces. NBER Working Paper 30633, National Bureau of Economic Research.
- Bartlett, R., A. Morse, R. Stanton, and N. Wallace (2022). Consumer-lending discrimination in the fintech era. *Journal of Financial Economics* 143(1), 30–56.
- Berg, T., V. Burg, A. Gombović, and M. Puri (2020). On the rise of fintechs: Credit scoring using digital footprints. *The Review of Financial Studies* 33(7), 2845–2897.
- Bhatta, A., V. Albiero, K. W. Bowyer, and M. C. King (2023). The gender gap in face recognition accuracy is a hairy problem. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) Workshops*, pp. 303–312.
- Blattner, L. and S. Nelson (2021). How costly is noise? data and disparities in consumer credit. *Working Paper*.
- Brock, J. M. and R. De Haas (2023). Discriminatory lending: Evidence from bankers in the lab. *American Economic Journal: Applied Economics* 15(2), 31–68.
- Brown, S. and K. Taylor (2014). Household finances and the ‘big five’ personality traits. *Journal of Economic Psychology* 45, 197–212.
- Celiktutan, O., F. Eyben, E. Sariyanidi, H. Gunes, and B. Schuller (2014). MAPTRAITS 2014: The first audio/visual mapping personality traits challenge – an introduction: Perceived personality and social dimensions. In *ICMI 2014: Proceedings of the 2014 International Conference on Multimodal Interaction*, Istanbul, Turkey, pp. 529–530. Association for Computing Machinery.
- Claes, P., D. K. Liberton, K. Daniels, K. M. Rosana, E. E. Quillen, L. N. Pearson, B. McEvoy, M. Bauchet, A. A. Zaidi, W. Yao, et al. (2014). Modeling 3d facial shape from dna. *PLOS Genetics* 10(3), e1004224.

- Connolly, J. J., E. J. Kavanagh, and C. Viswesvaran (2007). The convergent validity between self and observer ratings of personality: A meta-analytic review. *International Journal of Selection and Assessment* 15(1), 110–117.
- Deng, J., J. Guo, N. Xue, and S. Zafeiriou (2019). Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4690–4699.
- Donnelly, G., R. Iyer, and R. T. Howell (2012). The big five personality traits, material values, and financial well-being of self-described money managers. *Journal of Economic Psychology* 33(6), 1129–1142.
- Duarte, J., S. Siegel, and L. Young (2012). Trust and credit: The role of appearance in peer-to-peer lending. *The Review of Financial Studies* 25(8), 2455–2484.
- Fang, L. and H. J. Munneke (2020). Gender equality in mortgage lending. *Real Estate Economics* 48(4), 957–1003.
- Fuster, A., P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther (2022). Predictably unequal? the effects of machine learning on credit markets. *The Journal of Finance* 77(1), 5–47.
- Gibbs, C., B. Guttman-Kenney, D. Lee, S. Nelson, H. W. van der Klaauw, and J. Wang (2025). Consumer credit reporting data. *Journal of Economic Literature* 63(2), 598–636.
- Goldsmith-Pinkham, P. and K. Shue (2023). The gender gap in housing returns. *The Journal of Finance* 78(2), 1097–1145.
- Goodman, L. (2024). Classic FICO versus VantageScore 4.0. Urban Institute Report.
- Guenzel, M., S. Kogan, M. Niessner, and K. Shue (2025). AI personality extraction from faces: Labor market implications. Technical Report SSRN 5089827, SSRN.
- Hanley, J. A. and B. J. McNeil (1982). The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology* 143(1), 29–36.
- He, K., X. Zhang, S. Ren, and J. Sun (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778.
- Iyer, R., A. I. Khwaja, E. F. P. Luttmer, and K. Shue (2016). Screening peers softly: Inferring the quality of small borrowers. *Management Science* 62(6), 1554–1577.
- Jaffee, D. M. and T. Russell (1976). Imperfect information, uncertainty, and credit rationing. *The Quarterly Journal of Economics* 90(4), 651–666.
- Kachur, A., K. Osin, Y. Davydov, S. Shutilov, A. Novokshonov, A. Dorokhov, V. Gorbunova, S. Budyko, M. Y. Osipov, A. Moskvichev, D. Kuzmin, and A. Sboev (2020). Assessing the big five personality traits using real-life static facial images. *Scientific Reports* 10, 8487.
- Ke, G., Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu (2017). Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, Volume 30, pp. 3146–3154.

- Kleinberg, J., H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan (2018). Human decisions and machine predictions. *The Quarterly Journal of Economics* 133(1), 237–293.
- Laufer, S. and A. Paciorek (2022). The effects of mortgage credit availability: Evidence from minimum credit score lending rules. *American Economic Journal: Economic Policy* 14(1), 240–276.
- Lee, D. and W. van der Klaauw (2010, November). An introduction to the FRBNY consumer credit panel. Staff Report 479, Federal Reserve Bank of New York.
- Lin, M., N. R. Prabhala, and S. Viswanathan (2013). Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending. *Management Science* 59(1), 17–35.
- Ludwig, J. and S. Mullainathan (2024). Machine learning as a tool for hypothesis generation. *The Quarterly Journal of Economics* 139(2), 751–827.
- Penton-Voak, I. S., N. Pound, A. C. Little, and D. I. Perrett (2006). Personality judgments from natural and composite facial images: More evidence for a ‘kernel of truth’ in social perception. *Social Cognition* 24(5), 607–640.
- Peterson, J. C., S. Uddenberg, T. L. Griffiths, A. Todorov, and J. W. Suchow (2022). Deep models of superficial face judgments. *Proceedings of the National Academy of Sciences* 119(17), e2115228119.
- Phelps, E. S. (1972). The statistical theory of racism and sexism. *The American Economic Review* 62(4), 659–661.
- Pope, D. G. and J. R. Sydnor (2011). What’s in a picture? evidence of discrimination from prosper.com. *Journal of Human Resources* 46(1), 53–92.
- Ravina, E. (2020). Love & loans: The effect of beauty and personal characteristics in credit markets. *Journal of Financial Economics* 137(2), 320–352.
- Sharpe, S. A. (1990). Asymmetric information, bank lending, and implicit contracts: A stylized model of customer relationships. *The Journal of Finance* 45(4), 1069–1087.
- Snyder, M. (1984). When belief creates reality. *Advances in Experimental Social Psychology* 18, 247–305.
- Stiglitz, J. E. and A. Weiss (1981). Credit rationing in markets with imperfect information. *The American Economic Review* 71(3), 393–410.
- The Economist (2016). Tests of character. *The Economist*. Finance & Economics section.
- Vallée, B. and Y. Zeng (2019). Marketplace lending: A new banking paradigm? *The Review of Financial Studies* 32(5), 1939–1982.
- Vukasović, T. and D. Bratko (2015). Heritability of personality: A meta-analysis of behavior genetic studies. *Psychological Bulletin* 141(4), 769–785.

Table 1: Data Description

This table provides a description of key variables used in our analysis. All data are from 2019. $P(\text{Sex} = \text{Male})$ and $P(\text{Ethnicity} = \text{White})$ are indicator variables for male gender and White ethnicity, respectively. Age is measured in years and Salary in U.S. dollars. Thin file indicates borrowers with one or fewer credit accounts at t-2. US school rank represents educational institution ranking. Vantage score is the traditional VantageScore credit score, while Facial credit score is constructed using our ML algorithm applied to facial features. We employ two delinquency definitions, each measured using traditional credit bureau data and their facial ML predictions. The first definition identifies borrowers with a 30+ days past-due episode in the 12 months preceding t, conditional on being current at t-24. The second definition identifies borrowers experiencing an increase in 30+ days past-due tradelines during the same window. Photo Perceived Big 5 are based on an end-to-end prediction system is trained on the ChaLearn First Impressions dataset.

	Obs	Mean	Std. Dev.	P10	Median	P90
$P(\text{Sex} = \text{Male})$	1,100,628	0.45	0.47	0.00	0.16	1.00
$P(\text{Ethnicity} = \text{White})$	1,100,628	0.73	0.34	0.08	0.92	1.00
Age	1,100,628	31.28	10.93	20.00	28.00	48.00
Salary (\$)	1,050,097	74828.09	44618.95	31720.52	66042.54	123050.57
Thin file (≤ 1 accounts, t-2)	978,374	0.18	0.38	0.00	0.00	1.00
US school rank	668,456	321.83	226.82	46.00	275.00	655.00
Vantage score	1,100,628	713.02	93.30	561.00	729.00	817.00
Facial credit score	1,100,628	704.92	32.51	662.34	707.46	744.21
$\mathbb{1}(\text{30d+ delinq past 12m \& current in t-2})$	878,662	0.07	0.25	0.00	0.00	0.00
Facial delinquency (30d+ delinq past 12m & current in t-2)	880,359	0.07	0.02	0.04	0.07	0.10
$\mathbb{1}(\text{Increase 30d+ delinq past 12m})$	966,885	0.07	0.26	0.00	0.00	0.00
Facial delinquency (Increase 30d+ delinq past 12m)	967,854	0.07	0.02	0.05	0.07	0.09
Agreeableness	1,070,008	0.57	0.06	0.49	0.58	0.65
Conscientiousness	1,070,008	0.56	0.09	0.44	0.56	0.67
Extraversion	1,070,008	0.50	0.09	0.38	0.50	0.61
Neuroticism	1,070,008	0.55	0.08	0.44	0.55	0.65
Openness	1,070,008	0.58	0.08	0.48	0.59	0.68

Table 2: Vantage Score Prediction

This table regresses Vantage scores on the facial credit score, log salary and U.S. undergraduate school rank. Panel A presents specifications including only the facial credit score and demographic fixed effects or age controls. Panel B adds controls for log salary and U.S. undergraduate school rank. Variables denoted with (Z) are standardized to have mean 0 and standard deviation 1. Sex $\times$ ethnicity fixed effects and age controls are included where indicated. *Top20–Bottom20* is the difference between the average predicted Vantage score of the top and bottom quintile of individuals based on the facial credit score, expressed in standard deviation units. Robust standard errors are reported in parentheses. Significance levels are indicated by $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Panel A

	Vantage score (Z)				
	(1)	(2)	(3)	(4)	(5)
Facial credit score (Z)				0.356*** (0.001)	0.199*** (0.001)
Sex $\times$ Ethnicity FE	Yes	No	Yes	No	Yes
Age + Age ²	No	Yes	Yes	No	Yes
R2	0.034	0.137	0.163	0.127	0.191
Observations	1,100,628	1,100,628	1,100,628	1,100,628	1,100,628
Top20-Bottom20				0.989	0.553

Panel B

	Vantage score (Z)					
	(1)	(2)	(3)	(4)	(5)	(6)
Facial credit score (Z)		0.345*** (0.001)			0.278*** (0.001)	0.181*** (0.001)
Log salary (Z)			0.289*** (0.001)		0.195*** (0.001)	0.128*** (0.001)
US school rank (Z)				-0.125*** (0.001)	-0.070*** (0.001)	-0.078*** (0.001)
No US school rank				-0.091*** (0.002)	-0.096*** (0.002)	-0.105*** (0.002)
Sex $\times$ Ethnicity FE	Yes	No	No	No	No	Yes
Age + Age ²	Yes	No	No	No	No	Yes
R2	0.149	0.118	0.085	0.012	0.160	0.197
Observations	1,050,097	1,050,097	1,050,097	1,050,097	1,050,097	1,050,097
Top20-Bottom20		0.946			0.762	0.495

Table 3: Vantage Score Prediction by Sex

This table regresses Vantage scores on the facial credit score, log salary and U.S. undergraduate school rank. Panel A (B) presents for the female-only (male-only) sample specifications including the facial credit score, age controls and controls for log salary and U.S. undergraduate school rank. Variables denoted with (Z) are standardized to have mean 0 and standard deviation 1. Sex $\times$ ethnicity fixed effects and age controls are included where indicated. *Top20-Bottom20* is the difference between the average predicted Vantage score of the top and bottom quintile of individuals based on the facial credit score, expressed in standard deviation units. Robust standard errors are reported in parentheses. Significance levels are indicated by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Panel A: Female-Only Sample

	Vantage score (Z)					
	(1)	(2)	(3)	(4)	(5)	(6)
Facial credit score (Z)		0.365*** (0.001)			0.305*** (0.001)	0.210*** (0.001)
Log salary (Z)			0.278*** (0.001)		0.181*** (0.001)	0.121*** (0.001)
US school rank (Z)				-0.127*** (0.002)	-0.074*** (0.002)	-0.082*** (0.002)
No US school rank				-0.105*** (0.003)	-0.101*** (0.003)	-0.108*** (0.003)
Ethnicity FE	Yes	No	No	No	No	Yes
Age + Age ²	Yes	No	No	No	No	Yes
R2	0.146	0.132	0.079	0.012	0.170	0.203
Observations	577,222	577,222	577,222	577,222	577,222	577,222
Top20-Bottom20		1.001			0.838	0.576

Panel B: Male-Only Sample

	Vantage score (Z)					
	(1)	(2)	(3)	(4)	(5)	(6)
Facial credit score (Z)		0.313*** (0.001)			0.241*** (0.001)	0.140*** (0.002)
Log salary (Z)			0.295*** (0.001)		0.212*** (0.001)	0.133*** (0.002)
US school rank (Z)				-0.120*** (0.002)	-0.066*** (0.002)	-0.074*** (0.002)
No US school rank				-0.070*** (0.003)	-0.088*** (0.003)	-0.101*** (0.003)
Ethnicity FE	Yes	No	No	No	No	Yes
Age + Age ²	Yes	No	No	No	No	Yes
R2	0.148	0.097	0.089	0.010	0.145	0.188
Observations	472,875	472,875	472,875	472,875	472,875	472,875
Top20-Bottom20		0.858			0.660	0.385

Table 4: Delinquencies Prediction

This table regresses Delinquencies on the facial delinquencies score, log salary, U.S. undergraduate school rank, and Vantage score. Panel A presents specifications including only the facial delinquencies score and demographic fixed effects or age controls. Panel B adds controls for log salary, U.S. undergraduate school rank, and lagged Vantage score. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30-days-or-more past-due episode on a tradeline at some point in the 12 months immediately preceding t (the window $[t - 12, t]$), conditional on being current at $t - 24$. Variables denoted with (Z) are standardized to have mean 0 and standard deviation 1. Sex $\times$ ethnicity fixed effects and age controls are included where indicated. *Top20-Bottom20* is the difference between the average predicted delinquencies score of the top and bottom quintile of individuals based on the facial delinquencies score. Robust standard errors are reported in parentheses. Significance levels are indicated by $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Panel A

	Delinquency				
	(1)	(2)	(3)	(4)	(5)
Facial delinquency (Z)				0.026*** (0.000)	0.022*** (0.000)
Sex $\times$ Ethnicity FE	Yes	No	Yes	No	Yes
Age + Age ²	No	Yes	Yes	No	Yes
R2	0.005	0.002	0.007	0.011	0.014
Observations	878,662	878,662	878,662	878,662	878,662
Top20-Bottom20				0.071	0.061
AUC	0.563	0.549	0.591	0.607	0.626

Panel B

	Delinquency					
	(1)	(2)	(3)	(4)	(5)	(6)
Facial delinquency (Z)		0.026*** (0.000)			0.025*** (0.000)	0.021*** (0.000)
Log salary (Z)			-0.004*** (0.000)		-0.002*** (0.000)	-0.005*** (0.000)
US school rank (Z)				0.011*** (0.000)	0.008*** (0.000)	0.007*** (0.000)
No US school rank				0.010*** (0.001)	0.008*** (0.001)	0.006*** (0.001)
Sex $\times$ Ethnicity FE	Yes	No	No	No	No	Yes
Age + Age ²	Yes	No	No	No	No	Yes
R2	0.007	0.010	0.000	0.002	0.011	0.015
Observations	853,641	853,641	853,641	853,641	853,641	853,641
Top20-Bottom20		0.071			0.068	0.058
AUC	0.591	0.607	0.520	0.544	0.611	0.629

Table 5: Delinquencies Prediction with Lagged Vantage

This table regresses Delinquencies on the facial delinquencies score, log salary, U.S. undergraduate school rank, and Vantage score. Panel A presents specifications including the facial delinquencies score, lagged Vantaged score and demographic fixed effects or age controls. Panel B adds controls for log salary, U.S. undergraduate school rank, and lagged Vantage score. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30-days-or-more past-due episode on a tradeline at some point in the 12 months immediately preceding t (the window $[t - 12, t]$), conditional on being current at $t - 24$. Variables denoted with (Z) are standardized to have mean 0 and standard deviation 1. Sex $\times$ ethnicity fixed effects and age controls are included where indicated. *Top20-Bottom20* is the difference between the average predicted delinquencies score of the top and bottom quintile of individuals based on the facial delinquencies score. Robust standard errors are reported in parentheses. Significance levels are indicated by $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Panel A

	Delinquency				
	(1)	(2)	(3)	(4)	(5)
Facial delinquency (Z)		0.026*** (0.000)		0.022*** (0.000)	0.014*** (0.000)
Vantage score (Z, t-2)			-0.041*** (0.000)	-0.038*** (0.000)	-0.056*** (0.000)
Sex $\times$ Ethnicity FE	Yes	No	No	No	Yes
Age + Age ²	Yes	No	No	No	Yes
R2	0.007	0.011	0.023	0.030	0.044
Observations	878,416	878,416	878,416	878,416	878,416
Top20-Bottom20		0.071		0.061	0.038
AUC	0.591	0.607	0.677	0.692	0.722

Panel B

	Delinquency						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Facial delinquency (Z)		0.026*** (0.000)				0.022*** (0.000)	0.014*** (0.000)
Log salary (Z)			-0.004*** (0.000)			0.011*** (0.000)	0.001** (0.000)
US school rank (Z)				0.011*** (0.000)		0.005*** (0.000)	0.003*** (0.000)
No US school rank				0.010*** (0.001)		0.007*** (0.001)	0.002*** (0.001)
Vantage score (Z, t-2)					-0.042*** (0.000)	-0.043*** (0.000)	-0.057*** (0.000)
Sex $\times$ Ethnicity FE	Yes	No	No	No	No	No	Yes
Age + Age ²	Yes	No	No	No	No	No	Yes
R2	0.007	0.010	0.000	0.002	0.024	0.034	0.045
Observations	853,406	853,406	853,406	853,406	853,406	853,406	853,406
Top20-Bottom20		0.071				0.060	0.037
AUC	0.591	0.607	0.520	0.544	0.677	0.699	0.724

Table 6: Delinquencies Prediction by Sex

This table regresses Delinquencies on the facial delinquencies score, log salary, U.S. undergraduate school rank, and Vantage score. Panel A (B) presents for the female-only (male-only) sample specifications including the facial delinquencies score, age controls and controls for log salary, U.S. undergraduate school rank, and lagged Vantage score. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30-days-or-more past-due episode on a tradeline at some point in the 12 months immediately preceding t (the window $[t - 12, t]$), conditional on being current at $t - 24$. Variables denoted with (Z) are standardized to have mean 0 and standard deviation 1. Sex $\times$ ethnicity fixed effects and age controls are included where indicated. *Top20–Bottom20* is the difference between the average predicted delinquencies score of the top and bottom quintile of individuals based on the facial delinquencies score. Robust standard errors are reported in parentheses. Significance levels are indicated by $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Panel A: Female-Only Sample

	Delinquency						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Facial delinquency (Z)		0.029*** (0.000)				0.024*** (0.000)	0.015*** (0.000)
Log salary (Z)			-0.003*** (0.000)			0.011*** (0.000)	0.000 (0.000)
US school rank (Z)				0.011*** (0.000)		0.005*** (0.000)	0.003*** (0.000)
No US school rank				0.012*** (0.001)		0.007*** (0.001)	0.003*** (0.001)
Vantage score (Z, t-2)					-0.044*** (0.000)	-0.044*** (0.000)	-0.059*** (0.000)
Ethnicity FE							
Age + Age ²	Yes	No	No	No	No	No	Yes
R2	0.007	0.012	0.000	0.002	0.024	0.034	0.047
Observations	463,680	463,680	463,680	463,680	463,680	463,680	463,680
Top20-Bottom20		0.078				0.064	0.042
AUC	0.588	0.612	0.512	0.543	0.671	0.694	0.720

Panel B: Male-Only Sample

	Delinquency						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Facial delinquency (Z)		0.021*** (0.000)				0.018*** (0.000)	0.011*** (0.000)
Log salary (Z)			-0.003*** (0.000)			0.011*** (0.000)	0.001** (0.000)
US school rank (Z)				0.010*** (0.000)		0.005*** (0.000)	0.004*** (0.000)
No US school rank				0.008*** (0.001)		0.006*** (0.001)	0.001 (0.001)
Vantage score (Z, t-2)					-0.040*** (0.000)	-0.043*** (0.000)	-0.055*** (0.000)
Ethnicity FE							
Age + Age ²	Yes	No	No	No	No	No	Yes
R2	0.006	0.007	0.000	0.001	0.024	0.032	0.043
Observations	389,726	389,726	389,726	389,726	389,726	389,726	389,726
Top20-Bottom20		0.057				0.050	0.029
AUC	0.585	0.593	0.520	0.544	0.683	0.703	0.726

Table 7: Thick versus Thin Files

This table splits the sample by credit file thickness. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30-days-or-more past-due episode on a tradeline at some point in the 12 months immediately preceding t (the window $[t - 12, t]$), conditional on being current at $t - 24$. In Panel A we regress Vantage score on the facial credit score, and in Panel B we regress *Delinquency* on the facial delinquencies score and lagged Vantage score. A *thick file* is defined as having two or more credit lines, and a *thin file* has one or zero trade lines as of $t - 2$. All specifications include sex $\times$ ethnicity fixed effects and, where indicated, age controls. In Panel B, lagged Vantage score (Z, $t-2$) is added where indicated. Variables denoted with (Z) are standardized to have mean 0 and standard deviation 1. *Top20-Bottom20* is the difference between the average predicted delinquencies score of the top and bottom quintiles based on the corresponding facial score within each subsample. Robust standard errors are reported in parentheses. Significance levels are indicated by $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Panel A

	Delinquency			
	Thick file		Thin file	
	(1)	(2)	(3)	(4)
Facial delinquency (Z)		0.025*** (0.000)		0.015*** (0.000)
Sex $\times$ Ethnicity FE	Yes	Yes	Yes	Yes
Age + Age ²	Yes	Yes	Yes	Yes
R2	0.005	0.013	0.009	0.015
Observations	714,513	714,513	163,903	163,903
Top20-Bottom20		0.067		0.043
AUC	0.569	0.610	0.629	0.668
E(Y)		0.076		0.038

Panel B

	Delinquency					
	Thick file			Thin file		
	(1)	(2)	(3)	(4)	(5)	(6)
Facial delinquency (Z)	0.025*** (0.000)		0.013*** (0.000)	0.015*** (0.000)		0.013*** (0.000)
Vantage score (Z, $t-2$)		-0.080*** (0.000)	-0.077*** (0.000)		-0.017*** (0.001)	-0.015*** (0.001)
Sex $\times$ Ethnicity FE	Yes	Yes	Yes	Yes	Yes	Yes
Age + Age ²	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.013	0.058	0.060	0.015	0.015	0.020
Observations	714,513	714,513	714,513	163,903	163,903	163,903
Top20-Bottom20	0.067		0.035	0.043		0.038
AUC	0.610	0.741	0.744	0.668	0.673	0.695
E(Y)		0.076			0.038	

Table 8: Thick versus Thin Files by Sex

This table splits the sample by credit file thickness. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30-days-or-more past-due episode on a tradeline at some point in the 12 months immediately preceding t (the window $[t - 12, t]$), conditional on being current at $t - 24$. Panel A (B) presents for the female-only (male-only) sample specifications where we regress *Delinquency* on the facial delinquencies score and lagged Vantage score. A *thick file* is defined as having two or more credit lines, and a *thin file* has one or zero trade lines as of $t - 2$. All specifications include sex $\times$ ethnicity fixed effects and, where indicated, age controls. In Panel B, lagged Vantage score (Z, t-2) is added where indicated. Variables denoted with (Z) are standardized to have mean 0 and standard deviation 1. *Top20-Bottom20* is the difference between the average predicted delinquencies score of the top and bottom quintiles based on the corresponding facial score within each subsample. Robust standard errors are reported in parentheses. Significance levels are indicated by $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Panel A: Female-Only Sample

	Delinquency					
	Thick file			Thin file		
	(1)	(2)	(3)	(4)	(5)	(6)
Facial delinquency (Z)	0.028*** (0.000)		0.014*** (0.000)	0.017*** (0.001)		0.015*** (0.001)
Vantage score (Z, t-2)		-0.084*** (0.001)	-0.080*** (0.001)		-0.019*** (0.001)	-0.016*** (0.001)
Ethnicity FE	Yes		Yes	Yes		Yes
Age + Age ²	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.014	0.059	0.061	0.017	0.016	0.022
Observations	386,513	386,513	386,513	91,611	91,611	91,611
Top20-Bottom20	0.075		0.039	0.048		0.042
AUC	0.613	0.735	0.739	0.670	0.673	0.698
E(Y)		0.082			0.041	

Panel B: Male-Only Sample

	Delinquency					
	Thick file			Thin file		
	(1)	(2)	(3)	(4)	(5)	(6)
Facial delinquency (Z)	0.020*** (0.000)		0.010*** (0.000)	0.012*** (0.001)		0.011*** (0.001)
Vantage score (Z, t-2)		-0.076*** (0.001)	-0.074*** (0.001)		-0.015*** (0.001)	-0.013*** (0.001)
Ethnicity FE	Yes	Yes	Yes	Yes	Yes	Yes
Age + Age ²	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.010	0.055	0.057	0.013	0.013	0.017
Observations	328,000	328,000	328,000	72,292	72,292	72,292
Top20-Bottom20	0.053		0.028	0.035		0.030
AUC	0.597	0.747	0.749	0.661	0.667	0.688
E(Y)		0.068			0.034	

Table 9: ROC Comparison for Thick versus Thin Files

Entries report the difference in true-positive rate (TPR) between the *Features+Vantage* model and the *Vantage-only* model at a fixed false-positive rate (FPR): $\Delta\text{TPR} = \text{TPR}_{\text{Features+Vantage}} - \text{TPR}_{\text{Vantage}}$. Positive values mean *Features+Vantage* attains higher sensitivity at the same FPR (local ROC dominance). Predictions are from logit models estimated on the same sample: (i) Vantage score (baseline) and (ii) facial delinquency score plus Vantage. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30-days-or-more past-due episode on a tradeline at some point in the 12 months immediately preceding t (the window $[t - 12, t]$), conditional on being current at $t - 24$. *Thin files* are borrowers with one or fewer trade lines as of $t - 2$; *Thick files* have two or more. Differences are estimated via binormal ROC regression with 100 bootstrap repetitions; standard errors in parentheses. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

	Full sample (1)	Thin files (2)	Thick files (3)
FPR = 0.10	0.031*** (0.001)	0.122*** (0.004)	-0.014*** (0.002)
FPR = 0.30	0.030*** (0.001)	0.127*** (0.004)	0.000 (0.001)
FPR = 0.50	0.019*** (0.001)	0.059*** (0.003)	0.008*** (0.001)
FPR = 0.70	0.009*** (0.001)	0.005** (0.002)	0.009*** (0.001)
FPR = 0.90	0.001*** (0.000)	-0.007*** (0.001)	0.005*** (0.000)
Observations	882442	165003	717439

Table 10: Vantage Score Prediction and Photo Perceived Big 5

This table regresses Vantage scores and facial credit scores against Photo Perceived Big 5 of individuals. Sex $\times$ ethnicity controls are included. Variables denoted with (Z) are standardized to have mean 0 and standard deviation 1. *Top20–Bottom20* is the difference between the average predicted Vantage score of the top and bottom quintile of individuals based on their combined Photo Perceived Big 5 scores in Panel A, and using the predicted facial credit score in Panel B. Robust standard errors are reported in parentheses. Significance levels are indicated by $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

	Vantage score (Z)	Facial credit score (Z)
	(1)	(2)
Agreeableness (Z)	0.268*** (0.008)	0.481*** (0.007)
Conscientiousness (Z)	0.178*** (0.006)	0.226*** (0.006)
Extraversion (Z)	0.116*** (0.010)	0.003 (0.009)
Neuroticism (Z)	-0.652*** (0.022)	-0.765*** (0.020)
Openness (Z)	0.203*** (0.011)	0.199*** (0.010)
Sex $\times$ Ethnicity FE	Yes	Yes
Age + Age ²	Yes	Yes
R ²	0.169	0.322
Observations	1,070,008	1,070,008
Top20-Bottom20	0.207	0.404

Table 11: Delinquency Prediction and Photo Perceived Big 5

This table regresses delinquency and facial delinquency against Photo Perceived Big 5 of individuals. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30-days-or-more past-due episode on a tradeline at some point in the 12 months immediately preceding t (the window $[t - 12, t]$), conditional on being current at $t - 24$. Sex $\times$ ethnicity controls are included. Variables denoted with (Z) are standardized to have mean 0 and standard deviation 1. *Top20-Bottom20* is the difference between the average predicted delinquency of the top and bottom quintile of individuals based on their combined Photo Perceived Big 5 scores. Robust standard errors are reported in parentheses. Significance levels are indicated by $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

	Delinquency (1)	Facial delinquency (2)
Agreeableness (Z)	-0.036*** (0.002)	-0.011*** (0.000)
Conscientiousness (Z)	-0.013*** (0.002)	-0.003*** (0.000)
Extraversion (Z)	-0.008*** (0.003)	-0.001** (0.000)
Neuroticism (Z)	0.066*** (0.007)	0.017*** (0.001)
Openness (Z)	-0.018*** (0.004)	-0.006*** (0.000)
Sex $\times$ Ethnicity FE	Yes	Yes
Age + Age ²	Yes	Yes
R ²	0.008	0.110
Observations	854,779	856,420
Top20-Bottom20	0.019	0.008
AUC	0.596	

Table 12: Thick versus Thin Files and Photo Perceived Big 5

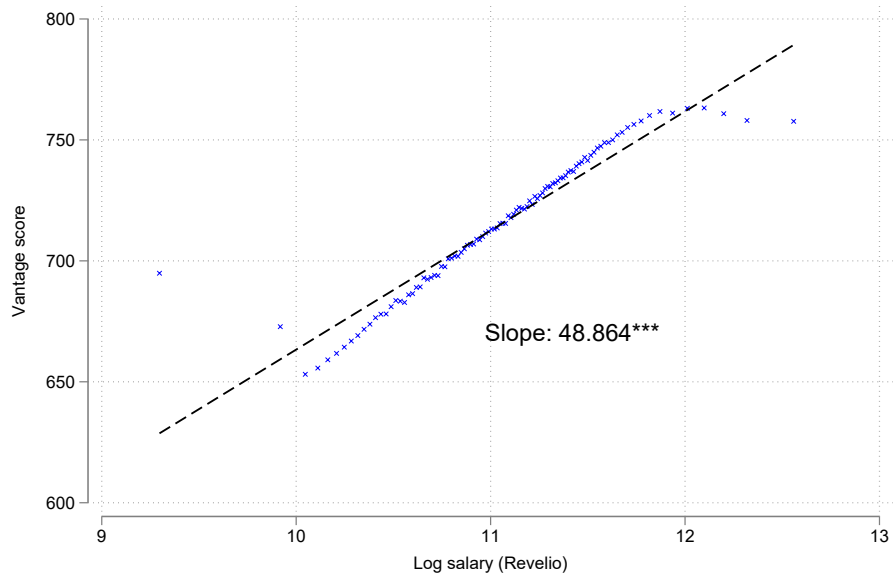
This table regresses delinquency and facial delinquency against Photo Perceived Big 5 of individuals for thick versus thin files. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30-days-or-more past-due episode on a tradeline at some point in the 12 months immediately preceding t (the window $[t - 12, t]$), conditional on being current at $t - 24$. A *thick file* is defined as having two or more credit lines, and a *thin file* has one or zero trade lines as of $t - 2$. Sex $\times$ ethnicity controls are included. Variables denoted with (Z) are standardized to have mean 0 and standard deviation 1. *Top20-Bottom20* is the difference between the average predicted delinquency of the top and bottom quintile of individuals based on their combined Photo Perceived Big 5 scores. Robust standard errors are reported in parentheses. Significance levels are indicated by $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

	Delinquency		Facial delinquency	
	Thick file	Thin file	Thick file	Thin file
	(1)	(2)	(3)	(4)
Openness (Z)	-0.017*** (0.004)	-0.018*** (0.006)	-0.005*** (0.000)	-0.015*** (0.001)
Conscientiousness (Z)	-0.016*** (0.002)	-0.003 (0.003)	-0.004*** (0.000)	-0.001*** (0.000)
Extraversion (Z)	-0.013*** (0.004)	-0.000 (0.006)	-0.001*** (0.000)	0.005*** (0.001)
Agreeableness (Z)	-0.036*** (0.003)	-0.025*** (0.004)	-0.010*** (0.000)	-0.017*** (0.000)
Neuroticism (Z)	0.072*** (0.008)	0.038*** (0.011)	0.016*** (0.001)	0.023*** (0.001)
Sex $\times$ Ethnicity FE	Yes	Yes	Yes	Yes
Age + Age ²	Yes	Yes	Yes	Yes
R ²	0.006	0.010	0.103	0.130
Observations	694,770	159,770	694,770	159,770
Top20-Bottom20	0.021	0.016	0.008	0.010
AUC	0.576	0.635		
E(Y)		0.076		0.038

Figure 1: Vantage Scores, Earnings, and Undergraduate Institution Rank

This figure presents binned scatter plots showing the intra-individual correlation of Vantage scores with (a) log salary and (b) U.S. undergraduate school rank, both obtained from Revelio Labs data. The solid lines represent fitted values from a linear regression.

a)



b)

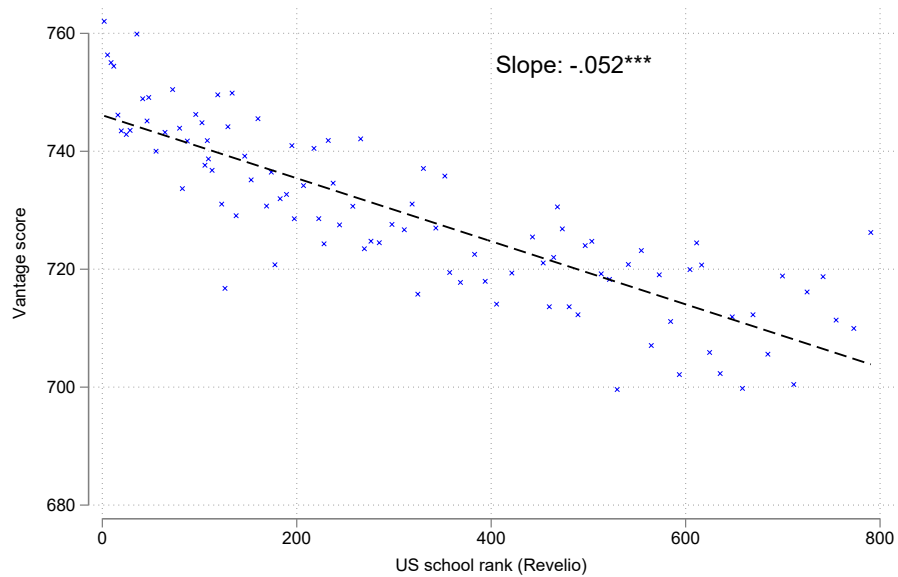


Figure 2: Number of Tradelines

This figure shows model performance across the borrowers' credit lifecycle, proxied by the number of tradelines on their credit report as of $t - 2$. It reports the Area Under Curve (AUC) for three prediction models: facial features only (red circles), Vantage score only (green squares), and facial features combined with Vantage (yellow diamonds), all estimated with demographic and age controls. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30-days-or-more past-due episode on a tradeline at some point in the 12 months immediately preceding t (the window $[t - 12, t]$), conditional on being current at $t - 24$.

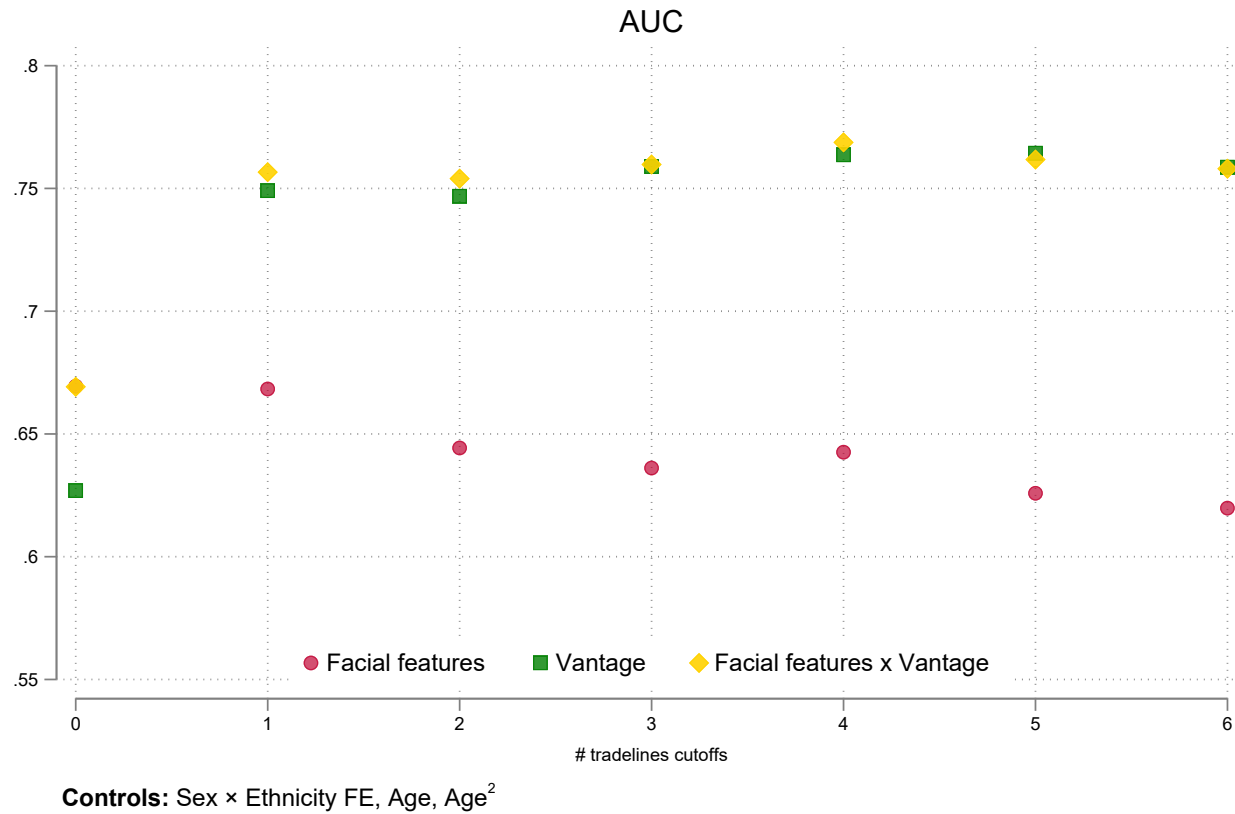


Figure 3: Borrower Age

This figure shows model performance across the borrowers' credit lifecycle, proxied by their age as of $t - 2$. It reports the Area Under Curve (AUC) for three prediction models: facial features only (red circles), Vantage score only (green squares), and facial features combined with Vantage (yellow diamonds), all estimated with demographic and age controls. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30-days-or-more past-due episode on a tradeline at some point in the 12 months immediately preceding t (the window $[t - 12, t]$), conditional on being current at $t - 24$.

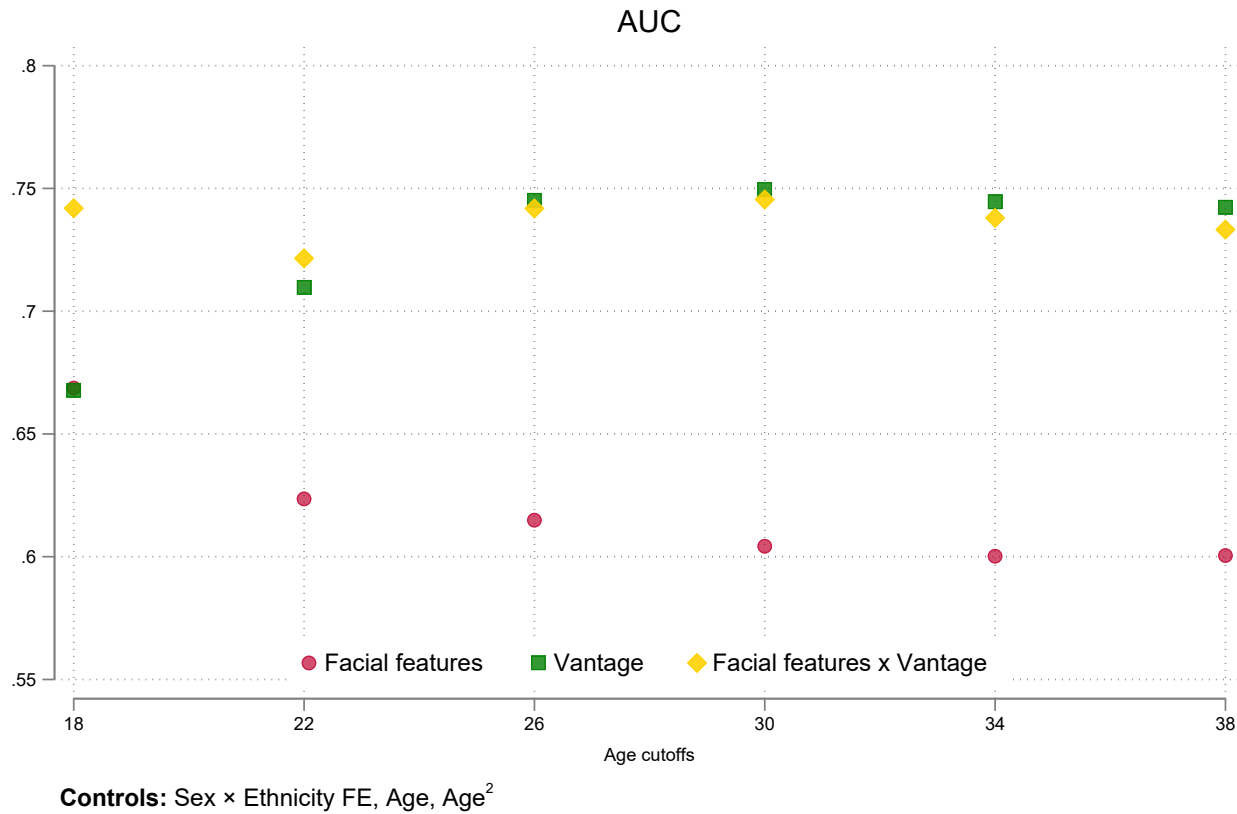


Figure 4: Borrower Income

This figure shows model performance across the borrowers' credit lifecycle, proxied by their income as of $t - 2$. It reports the Area Under Curve (AUC) for three prediction models: facial features only (red circles), Vantage score only (green squares), and facial features combined with Vantage (yellow diamonds), all estimated with demographic and age controls. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30-days-or-more past-due episode on a tradeline at some point in the 12 months immediately preceding t (the window $[t - 12, t]$), conditional on being current at $t - 24$.

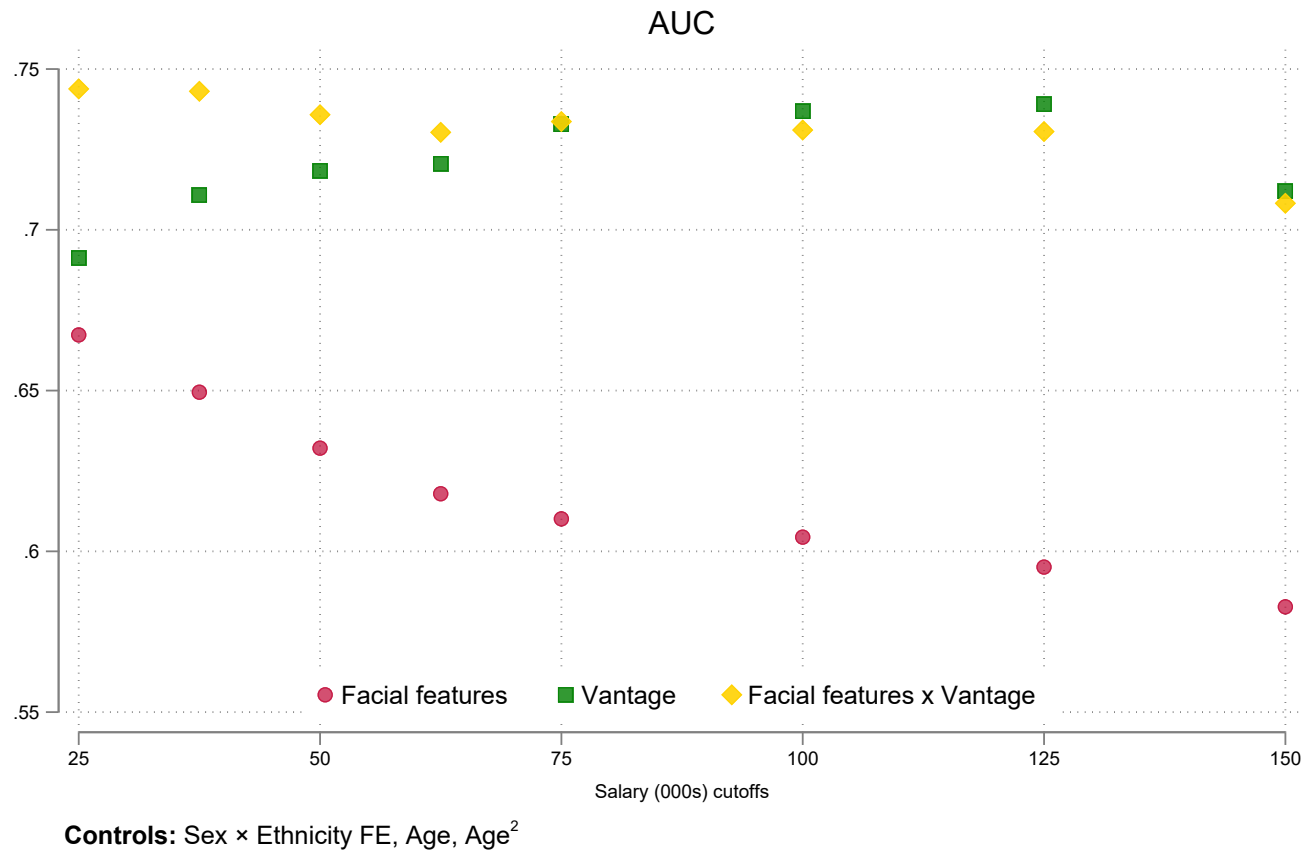
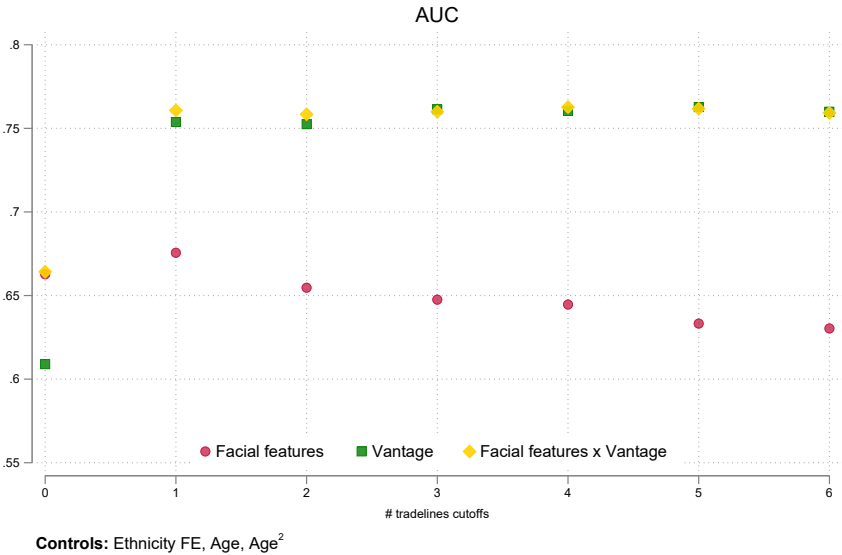
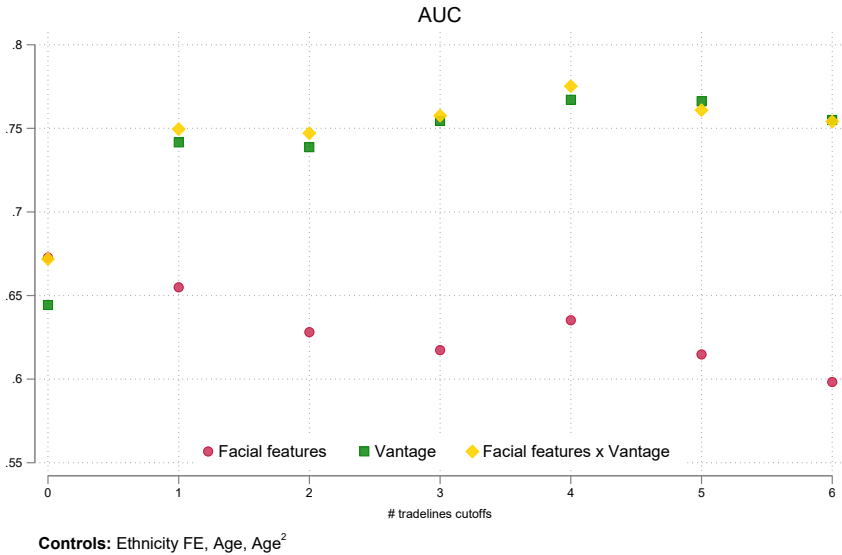


Figure 5: Number of Tradelines by Sex

This figure shows model performance across the borrowers' credit lifecycle, proxied by the number of tradelines on their credit report as of $t - 2$. It reports the Area Under Curve (AUC) for three prediction models: facial features only (red circles), Vantage score only (green squares), and facial features combined with Vantage (yellow diamonds), all estimated with demographic and age controls. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30-days-or-more past-due episode on a tradeline at some point in the 12 months immediately preceding t (the window $[t - 12, t]$), conditional on being current at $t - 24$. Panel (a) reports the AUCs for the female-only sample, and Panel (b) for the male-only sample.



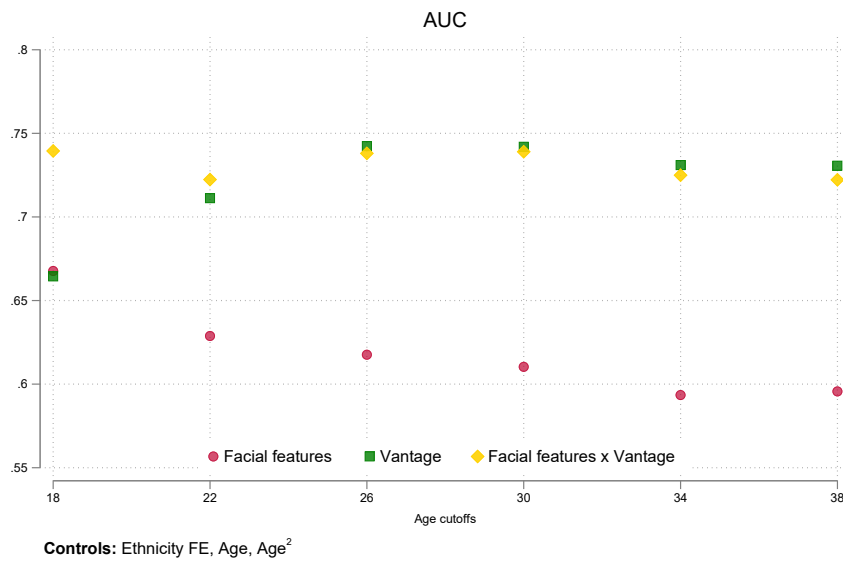
(a) Female-Only Sample



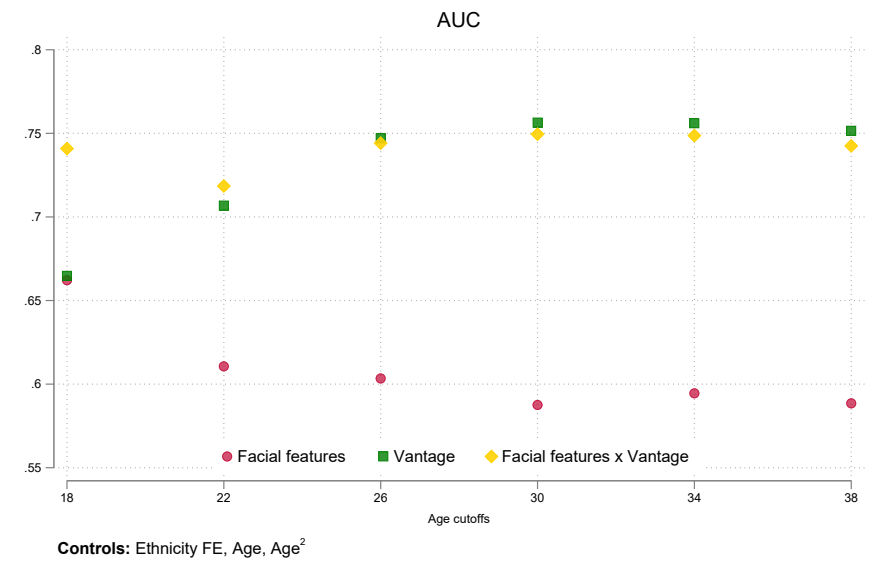
(b) Male-Only Sample

Figure 6: Borrower Age by Sex

This figure shows model performance across the borrowers' credit lifecycle, proxied by their age as of $t - 2$. It reports the Area Under Curve (AUC) for three prediction models: facial features only (red circles), Vantage score only (green squares), and facial features combined with Vantage (yellow diamonds), all estimated with demographic and age controls. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30-days-or-more past-due episode on a tradeline at some point in the 12 months immediately preceding t (the window $[t - 12, t]$), conditional on being current at $t - 24$. Panel (a) reports the AUCs for the female-only sample, and Panel (b) for the male-only sample.



(a) Female-Only Sample



(b) Male-Only Sample

Figure 7: Borrower Income by Sex

This figure shows model performance across the borrowers' credit lifecycle, proxied by their income as of $t - 2$. It reports the Area Under Curve (AUC) for three prediction models: facial features only (red circles), Vantage score only (green squares), and facial features combined with Vantage (yellow diamonds), all estimated with demographic and age controls. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30-days-or-more past-due episode on a tradeline at some point in the 12 months immediately preceding t (the window $[t - 12, t]$), conditional on being current at $t - 24$. Panel (a) reports the AUCs for the female-only sample, and Panel (b) for the male-only sample.

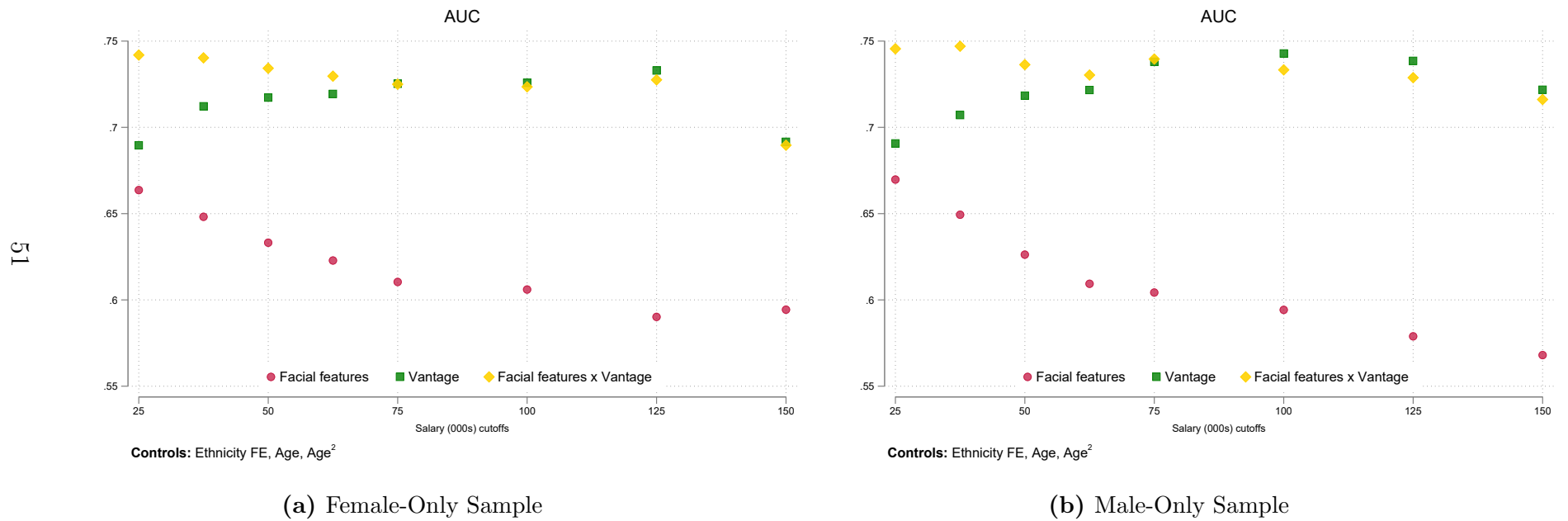
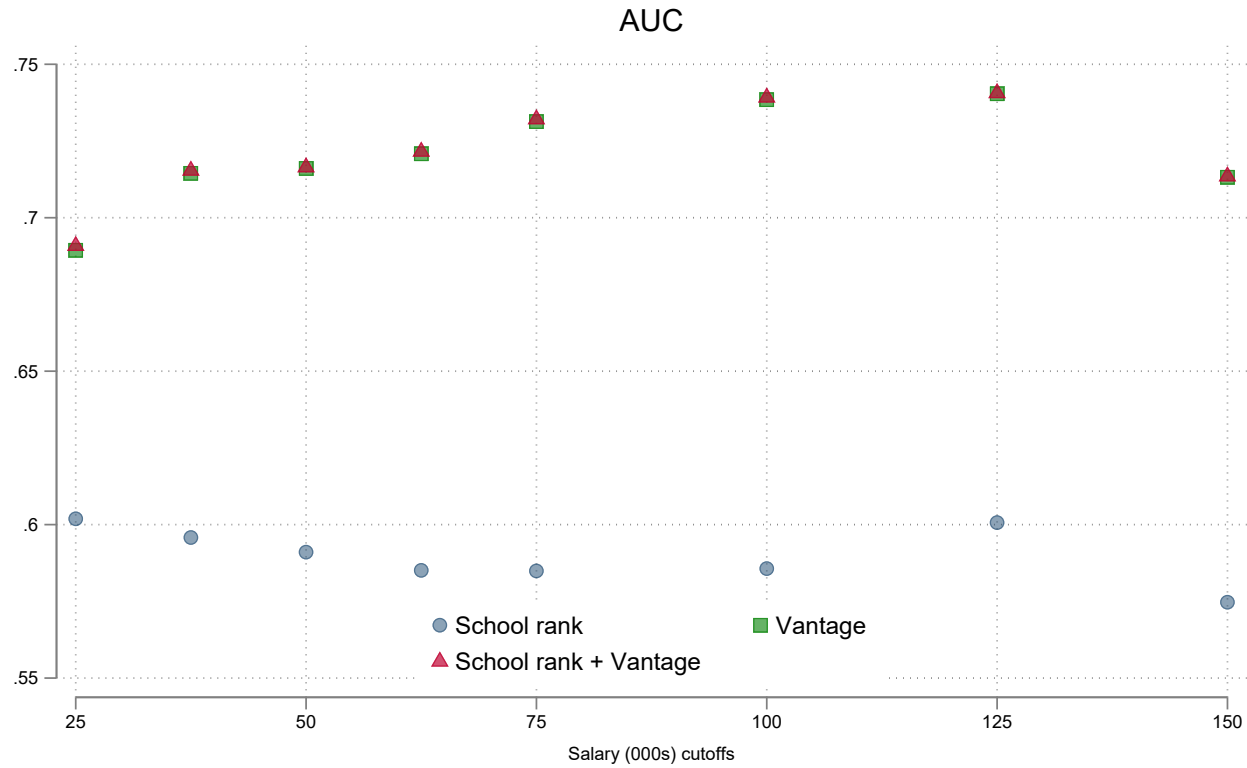


Figure 8: Borrower Income using School Rank

This plot shows the disparities in model performance across borrower income. It reports the Area Under Curve (AUC) for three prediction models: school rank (blue circles), Vantage score only (green squares), and school rank combined with Vantage (red triangles), all estimated with demographic and age controls. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30-days-or-more past-due episode on a tradeline at some point in the 12 months immediately preceding t (the window $[t - 12, t]$), conditional on being current at $t - 24$.



Controls: Sex $\times$ Ethnicity FE, Age, Age²

Figure 9: ROC Comparison

This plot shows Receiver Operating Characteristic (ROC) curve to measure the performance of the Vantage (blue) versus the Vantage combined with facial features (red). The ROC shows the trade-off between the true-positive rate (sensitivity) and the false-positive rate (1-specificity) at various decision thresholds of the two models. The ROC is always above the reference 45 degree line, with a more top-left curve indicating better performance in distinguishing true positives and true negatives. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30-days-or-more past-due episode on a tradeline at some point in the 12 months immediately preceding t (the window $[t - 12, t]$), conditional on being current at $t - 24$.

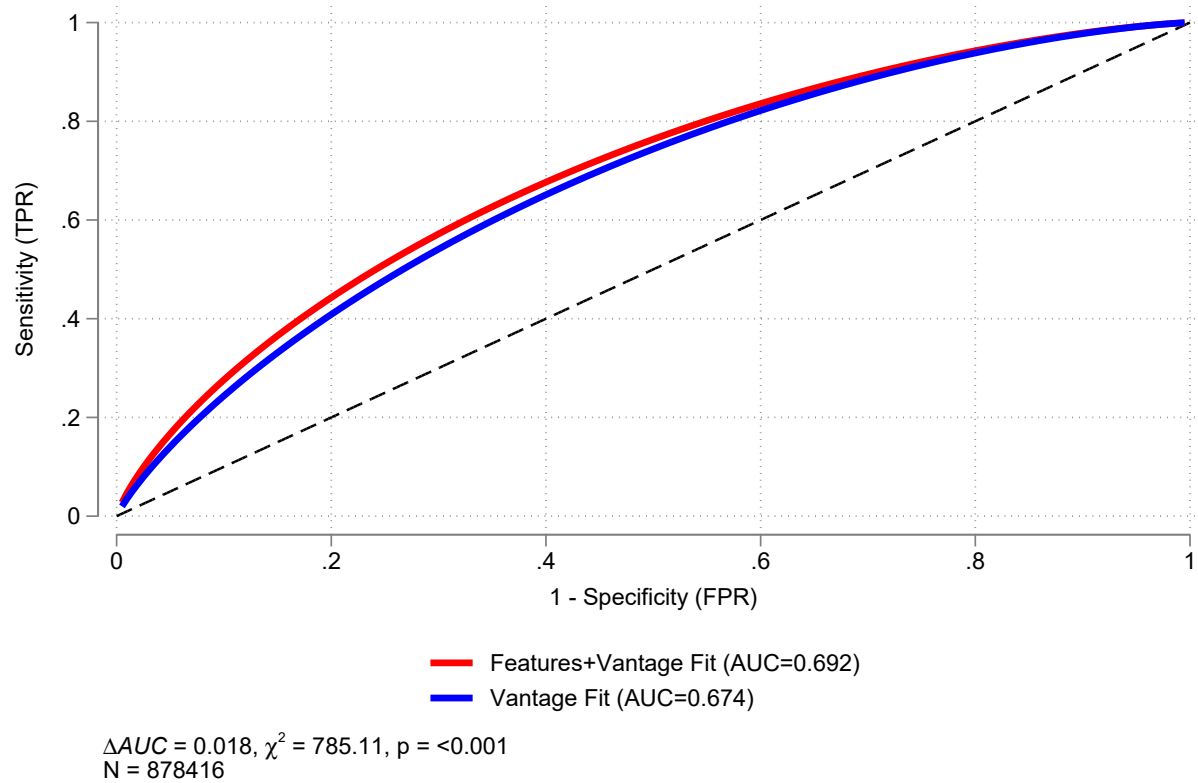
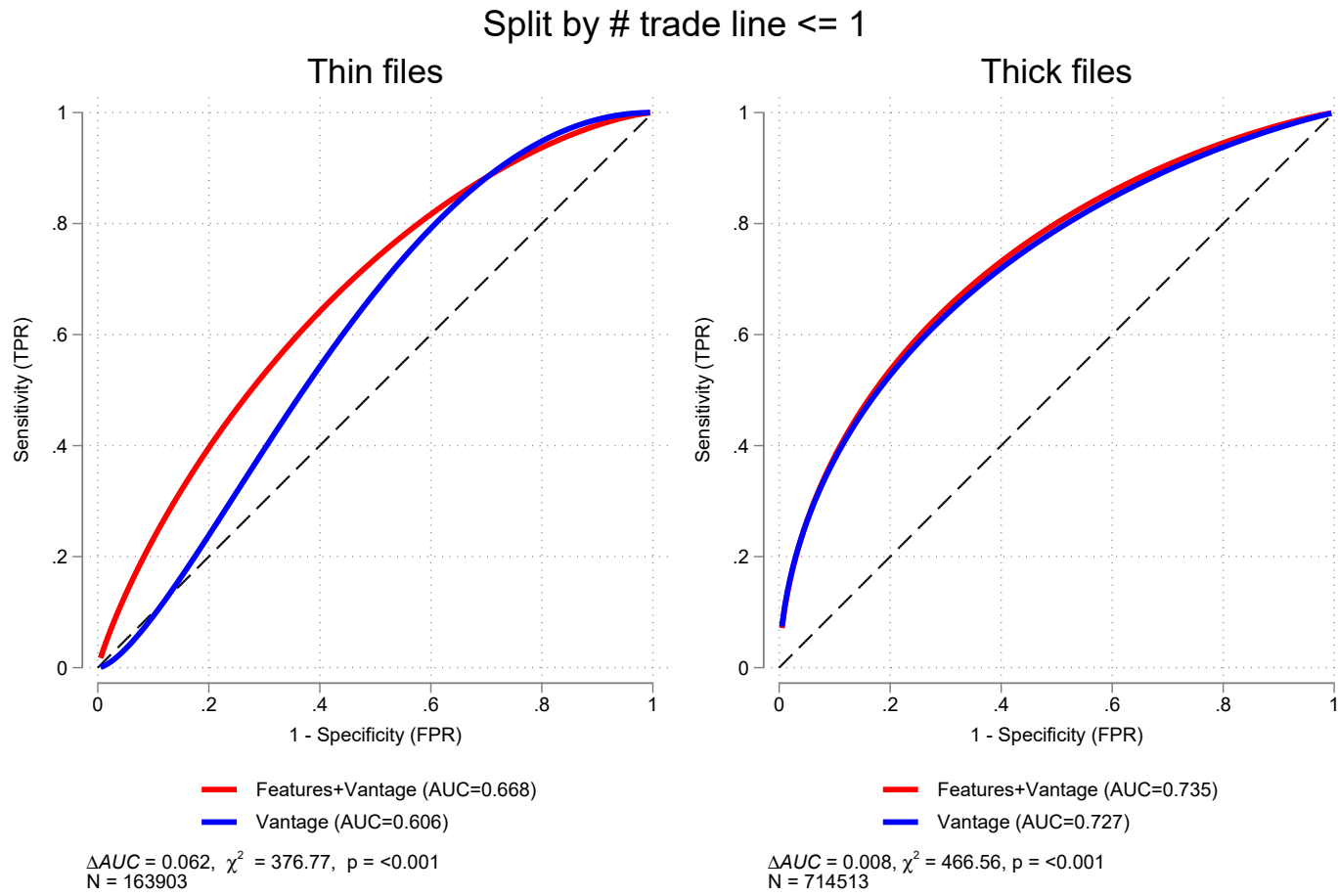


Figure 10: ROC Comparison for Thick versus Thin Files

This plot shows Receiver Operating Characteristic (ROC) curve to measure the performance of the Vantage (blue) versus the Vantage combined with facial features (red). The ROC shows the trade-off between the true-positive rate (sensitivity) and the false-positive rate (1-specificity) at various decision thresholds of the two models. The ROC is always above the reference 45 degree line, with a more top-left curve indicating better performance in distinguishing true positives and true negatives. We plot the ROC separately for thin versus thick files. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30-days-or-more past-due episode on a tradeline at some point in the 12 months immediately preceding t (the window $[t - 12, t]$), conditional on being current at $t - 24$.



Online Appendix FACING DEFAULT?

Pranav Garg, Marius Guenzel, Shimon Kogan, Marina Niessner, Kelly Shue

A. DATA CONSTRUCTION

A.1 CLEANING REVELIO DATA

The unique identifier in Revelio is `user_id`. Each individual, regardless of relocation across states, is assigned a single `user_id` in the data. There are three parent directories in Revelio: Profiles, Positions, and Education. The latter two are individual-by-time panel data, recording the employment and academic histories of individuals. The former is a snapshot of all LinkedIn users in the Revelio dataset. Our latest vintage of Revelio data is 2023-10-30.

First, we filter `user_id`'s into panels labeled by the 51 U.S. states.¹

1. We retain all U.S.-based profiles, positions and education degrees. Given that we are matching to the panel of registered voters in the U.S., this restriction is satisfactory since a majority of individuals are either U.S. natives or will need to be naturalized to show up in the L2 data. However, this restriction does preclude data on undergraduate degrees for immigrants who studied abroad and were later naturalized, which hinders our ability to calculate their age.²
2. We drop all observations with a missing location or name. We extract the state from the location string variable.
3. We drop all profiles with less than 25 connections. Individuals with less than 25 connections may not be legitimate LinkedIn users, and often have missing or incomplete information. This drops around 25% of the sample.
4. For both positions and education panels, we drop a `user_id` if:
 - Starting date is missing.
 - Period-of-stay is less than 90 days.
 - Academic institution name or employer name is missing.

From the start date of the first undergraduate degree, we calculate an estimate of the `user_id`'s birth year as `start_year - 18`.

¹If a `user_id` relocated from Texas to Michigan in March 2012, then they will be assigned to the Texas panel prior to March 2012, and to the Michigan panel after. We create state panels to account for the state-level data files provided by L2.

²The number of such cases could be too low to matter in the L2 data.

5. While cleaning the individual names associated with a `user_id`, we employ the following steps:

- Trim all white spaces.
- Drop everything within `()` or `[]` or `{}` or `"`. For example, "frank (real estate) greene".
- Drop everything after and including `,`. For example, "richie jackson, PhD".
- Remove the following symbols: `[# ! ? * ~ & @ \ / ^ % = $ > < ; : % +]`
- Remove a "junior, senior, jr, sr" from the start or the end of the string. For example, "jr jane doe".
- Drop single letters surrounded by spaces or at the start or the end.
- Drop names starting with or containing "the". For example, "the pc hero".
- Drop all non-Latin alphabet strings.
- Drop if < 2 distinct words in the name, count as one if hyphenated.

A.2 MERGING REVELIO AND L2

First, we do state-wise cleaning of both the L2 and Revelio datasets to filter the individuals suitable for a merge. Next, we attempt a fuzzy merge.

A.2.1 PRE-MERGE CLEANING

The unique identifier in the L2 data is `l Alvoterid`. Each individual voter is assigned a unique `l Alvoterid` by state, so that `l Alvoterid` is unique by person-state and not by person alone. A `l Alvoterid` is never re-assigned. Every L2 data file is provided as a state-wise cross-section of all registered voters for a given snapshot. For instance, this is a file for all registered voters in California in 2019-08-01.

1. From the L2 data, for every `l Alvoterid`, we extract information fields for the full name and DOB. We assign the state that the `l Alvoterid` was in based on the state-wise panel they are present in and the date of presence as the vintage date of the panel. We append all annual vintages for every state, with dates ranging from 2017 to 2024. This creates a single repeated cross-section for each state, with unique observations identifiable by .
2. For each state-level data file for Revelio and L2, we retain all observations with:
 - Non-missing name.
 - Non-missing start & end dates for every position or academic degree in that state.
 - Non-missing birth year.

These aforementioned steps drop no additional data since Revelio was already cleaned by name and the date ranges of state presence. Around 39-45% observations in Revelio have non-missing birth years, so I drop approximately 55-61% rows from the final condition.

A.2.2 FUZZY MERGE

The merge occurs state-by-state. For each state, the fuzzy merge occurs in three steps. First, retain all unique ID-name pairs in each state and conduct a fuzzy merge on the names. Next, we merge birth year and start-end dates to conduct additional rounds of cleaning. Finally, on the merged sample, we conduct a second round of cleaning for every merged pair to exclude false positives. Below, we detail all steps involved in the fuzzy merge procedure.

1. We retain all (laltoterid, L2 name) and (user_id, Revelio name) pairs. If there exist multiple names by laltoterid, we keep the longest string. For example, if “cecilia marie sanchez” or “cecilia m sanchez” are assigned the same laltoterid, I keep the former.
2. I conduct a fuzzy merge between the L2 and Revelio names based on a Weighted Jaccard Similarity algorithm for set similarity using the R package [fedmatch](#). This is one of the most commonly used algorithms for fuzzy matching, and is particularly useful in scenarios where elements can have varying importance (hence the “weighted”). The basic idea behind the algorithm is as follows:

- First, every string (name) is tokenized into sets of elements. A token for the core function `wgt_jaccard_distance` relies on unigrams (individual words). For example, “Sam Wutterberry Smith” is broken into 3-tokens: “Sam”, “Wutterberry”, and “Smith”.
- For every token, a weight is pre-assigned based on some external criteria such as frequency, importance, or relevance. These weights are not inherent to the token itself but are assigned depending on the context or some predefined rules. Let us say the weights are: $w(Sam) = 1$, $w(Wutterberry) = 4$, $w(Smith) = 1$.
- The weighted intersection is the sum of the minimum weights of tokens that appear in both sets. Let’s say the two strings are represented by token sets A and B, with weights $w_A(x)$ and $w_B(x)$ for every token x in string names A and B, respectively.³ The weighted intersection is:

$$I = \sum_{x \in A \cap B} \min(w_A(x), w_B(x))$$

- The weighted union is the sum of the maximum weights of tokens in both sets:

$$U = \sum_{x \in A \cup B} \max(w_A(x), w_B(x))$$

- Finally, the Weighted Jaccard Similarity is defined as the ratio $\frac{I}{U}$, and ranges between 0 (no similarity) and 1 (identical sets).
- As an example, let us illustrate this using string A as “cat” and string B as “bat”, with tokenization set as 1 letter.
String A: “cat”, weights for characters: c:1, a:1, t:2.

³Usually, $w_A(x) = w_B(x)$ for every unique token x , but I am providing a more general notation.

String B: “bat”, weights for characters: b:2, a:1, t:1.
Intersection weights = 1 (for “a”) + 1 (for “t”) = 2.
Union weights = 1 (for “c”) + 2 (for “b”) + 1 (for “a”) + 2 (for “t”) = 6.
Weighted similarity = $2/6 = 0.33$.

3. We set the threshold for weighted similarity (`max_dist`) as 0.1. This is fairly restrictive to eliminate spurious matches but fairly accommodative to pick up several nearly-good matches. We set 0.1 after several round of trial-and-error on smaller subset of randomly selected names.
4. Next, for every matched L2-Revelio name pair, we retain pairs that further satisfy the following criteria:
 - If date range is within ± 365 days of presence in state in L2 and Revelio.
 - If birth year is within ± 2 years in L2 and Revelio.
5. We sort data based on a descending similarly score for each `user_id` within each state, and keep the `l2voterid` with the highest score. This step ensures we have at most one match for every `user_id` in a state, but since we am looping over states, we can have the same individual match with multiple `l2voterid` across states.
6. Finally, across the sample of matched L2-Revelio individuals, we run a second round of string cleaning (this step was ChatGPT assisted) that is intended to remove almost all false positives from the sample. The following steps are employed on Python:
 - We retain all matches that have at most one spelling mistake in the name (per every 7 grams/letters in the name).
 - We retain all matches that have middle names missing or abridged.
 - Apart from the two exceptions mentioned above, we categorize all cases that are not exact matches (upto these two exception) as non-matched.
 - To check if the exact string match procedure is dropping too many correct matches, we conduct a phonetic match using the `metaphone` package. This ascertains a match between pre-defined pronunciations of various tokens, which may constitute a non-match by comparing string distance via the Jaccard algorithm. For example, `user_id` 4050087 for individual “Rasheedah Hasan” and `l2voterid` LALNY104566205 for individual “Rasheeda Hassan” were categorized as a non-match based on the exact string match procedure but categorized as a phonetic match.
 - Additionally, for each matched L2-Revelio individual, we calculate string matching scores based on other leading algorithms: `levenshtein_normalized_similarity`, `cosine_normalized_similarity`, `bag_normalized_similarity`, `mra_normalized_similarity`. We consult these scores for a benchmark on merge quality, but not for any restriction conditions.

B. CHALEARN PHOTO PERCEIVED BIG 5 PERFORMANCE

The Photo Perceived Big 5 are built for our dataset of 1.1 million individuals based on the prediction system trained on the ChaLearn First Impressions dataset.

Model training for our ChaLearn data uses the AdamW optimizer (learning rate 0.001, cosine annealing schedule) and Huber loss ($\delta = 0.1$) for robustness against outliers. Mixup augmentation (50% probability) and early stopping (15-epoch patience) enhance generalization and stability. Feature scaling is applied exclusively to training data to prevent leakage. Twelve configurations were trained combining engines and demographic variants. We try three different engines: InceptionResNetV1 (pretrained on CASIA-WebFace), ResNet-50, and Vision Transformers (ViT, Swin). We employ three layer sizes ($512 \rightarrow 256 \rightarrow 128 \rightarrow 64$) in the main branch, with a second branch takes demographic encodings as inputs. The two branches are fused in a final layer that outputs five normalized OCEAN scores. The best-performing model – InceptionResnetV1 pretrained on CASIA-WebFace without demographics – achieved $R^2 = 0.326$ and MAE = 0.096.

The table below reports the prediction performance of each of the five personalities measures. The correlations (Pearson) are with respect to the ground truth (i.e., the true personality scores from the ChaLearn data), specifically on the test data (2000 images from the videos).

For Conscientiousness and Extraversion, we observe a strong positive correlation, so the predictions are highly reliable. For Openness and Neuroticism, the correlation is moderate-to-strong, while for Agreeableness it is weaker, as this was the hardest trait for the model to learn.

Table B1: Performance metrics by personality dimension

Trait	R^2	MAE	Correlation
Conscientiousness	0.385	0.095	0.621
Extraversion	0.366	0.096	0.605
Neuroticism	0.335	0.099	0.578
Openness	0.307	0.096	0.554
Agreeableness	0.236	0.093	0.486

C. ROBUSTNESS: ALTERNATE DEFINITION OF DELINQUENCY

Table C1: Delinquencies Prediction

This table regresses Delinquencies on the facial delinquencies score, log salary, U.S. undergraduate school rank, and Vantage score. Panel A presents specifications including only the facial delinquencies score and demographic fixed effects or age controls. Panel B adds controls for log salary, U.S. undergraduate school rank, and lagged Vantage score. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences an increase in the number of tradelines that are 30-days-or-more past-due in the 12 months immediately preceding t (the window $[t - 12, t]$). Variables denoted with (Z) are standardized to have mean 0 and standard deviation 1. Sex $\times$ ethnicity fixed effects and age controls are included where indicated. *Top20-Bottom20* is the difference between the average predicted delinquencies score of the top and bottom quintile of individuals based on the facial delinquencies score. Robust standard errors are reported in parentheses. Significance levels are indicated by $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Panel A

	Delinquency				
	(1)	(2)	(3)	(4)	(5)
Facial delinquency (Z)		0.025*** (0.000)		0.020*** (0.000)	0.013*** (0.000)
Vantage score (Z, t-2)			-0.037*** (0.000)	-0.034*** (0.000)	-0.045*** (0.000)
Sex $\times$ Ethnicity FE	Yes	No	No	No	Yes
Age + Age ²	Yes	No	No	No	Yes
R2	0.007	0.009	0.020	0.026	0.037
Observations	966,625	966,625	966,625	966,625	966,625
Top20-Bottom20		0.068		0.055	0.036
AUC	0.590	0.601	0.669	0.683	0.710

Panel B

	Delinquency						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Facial delinquency (Z)		0.025*** (0.000)				0.020*** (0.000)	0.013*** (0.000)
Log salary (Z)			-0.004*** (0.000)			0.009*** (0.000)	0.000 (0.000)
US school rank (Z)				0.011*** (0.000)		0.005*** (0.000)	0.004*** (0.000)
No US school rank				0.011*** (0.001)		0.006*** (0.001)	0.002*** (0.001)
Vantage score (Z, t-2)					-0.038*** (0.000)	-0.038*** (0.000)	-0.045*** (0.000)
Sex $\times$ Ethnicity FE	Yes	No	No	No	No	No	Yes
Age + Age ²	Yes	No	No	No	No	No	Yes
R2	0.007	0.009	0.000	0.002	0.021	0.029	0.037
Observations	940,494	940,494	940,494	940,494	940,494	940,494	940,494
Top20-Bottom20		0.067				0.054	0.035
AUC	0.590	0.601	0.518	0.543	0.669	0.689	0.711

Table C2: Thick versus Thin Files

This table splits the sample by credit file thickness. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences an increase in the number of tradelines that are 30-days-or-more past-due in the 12 months immediately preceding t (the window $[t - 12, t]$). In Panel A we regress Vantage score on the facial credit score, and in Panel B we regress *Delinquency* on the facial delinquencies score and lagged Vantage score. A *thick file* is defined as having two or more credit lines, and a *thin file* has one or zero trade lines as of $t - 2$. All specifications include sex $\times$ ethnicity fixed effects and, where indicated, age controls. In Panel B, lagged Vantage score ($Z, t-2$) is added where indicated. Variables denoted with (Z) are standardized to have mean 0 and standard deviation 1. *Top20-Bottom20* is the difference between the average predicted delinquencies score of the top and bottom quintiles based on the corresponding facial score within each subsample. Robust standard errors are reported in parentheses. Significance levels are indicated by $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Panel A

	Delinquency			
	Thick file		Thin file	
	(1)	(2)	(3)	(4)
Facial delinquency (Z)		0.023*** (0.000)		0.015*** (0.000)
Sex $\times$ Ethnicity FE	Yes	Yes	Yes	Yes
Age + Age ²	Yes	Yes	Yes	Yes
R2	0.005	0.011	0.009	0.015
Observations	800,592	800,592	166,033	166,033
Top20-Bottom20		0.061		0.042
AUC	0.569	0.604	0.628	0.666
E(Y)		0.079		0.038

Panel B

	Delinquency					
	Thick file			Thin file		
	(1)	(2)	(3)	(4)	(5)	(6)
Facial delinquency (Z)	0.023*** (0.000)		0.012*** (0.000)	0.015*** (0.000)		0.013*** (0.000)
Vantage score (Z, t-2)		-0.059*** (0.000)	-0.056*** (0.000)		-0.017*** (0.001)	-0.014*** (0.001)
Sex $\times$ Ethnicity FE	Yes	Yes	Yes	Yes	Yes	Yes
Age + Age ²	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.011	0.041	0.043	0.015	0.015	0.019
Observations	800,592	800,592	800,592	166,033	166,033	166,033
Top20-Bottom20	0.061		0.033	0.042		0.037
AUC	0.604	0.721	0.723	0.666	0.671	0.693
E(Y)		0.079			0.038	

Table C3: Delinquency Prediction and Photo Perceived Big 5

This table regresses delinquency and facial delinquency against Photo Perceived Big 5 of individuals. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences an increase in the number of tradelines that are 30-days-or-more past-due in the 12 months immediately preceding t (the window $[t - 12, t]$). Sex $\times$ ethnicity controls are included. Variables denoted with (Z) are standardized to have mean 0 and standard deviation 1. *Top20-Bottom20* is the difference between the average predicted delinquency of the top and bottom quintile of individuals based on their combined Photo Big 5 scores. Robust standard errors are reported in parentheses. Significance levels are indicated by $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

	Delinquency (1)	Facial delinquency (2)
Agreeableness (Z)	-0.036*** (0.002)	-0.011*** (0.000)
Conscientiousness (Z)	-0.013*** (0.002)	-0.003*** (0.000)
Extraversion (Z)	-0.008*** (0.003)	-0.001** (0.000)
Neuroticism (Z)	0.066*** (0.007)	0.017*** (0.001)
Openness (Z)	-0.018*** (0.004)	-0.006*** (0.000)
Sex $\times$ Ethnicity FE	Yes	Yes
Age + Age ²	Yes	Yes
R ²	0.008	0.110
Observations	854,779	856,420
Top20-Bottom20	0.019	0.008
AUC	0.596	

Table C4: Thick versus Thin Files and Photo Perceived Big 5

This table regresses delinquency and facial delinquency against Photo Perceived Big 5 of individuals for thick versus thin files. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences an increase in the number of tradelines that are 30-days-or-more past-due in the 12 months immediately preceding t (the window $[t - 12, t]$). A *thick file* is defined as having two or more credit lines, and a *thin file* has one or zero trade lines as of $t - 2$. Sex $\times$ ethnicity controls are included. Variables denoted with (Z) are standardized to have mean 0 and standard deviation 1. *Top20-Bottom20* is the difference between the average predicted delinquency of the top and bottom quintile of individuals based on their combined Photo Big 5 scores. Robust standard errors are reported in parentheses. Significance levels are indicated by $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

	Delinquency		Facial delinquency	
	Thick file	Thin file	Thick file	Thin file
	(1)	(2)	(3)	(4)
Openness (Z)	-0.014*** (0.004)	-0.018*** (0.006)	-0.004*** (0.000)	-0.013*** (0.001)
Conscientiousness (Z)	-0.015*** (0.002)	-0.003 (0.003)	-0.003*** (0.000)	-0.001*** (0.000)
Extraversion (Z)	-0.013*** (0.003)	0.000 (0.006)	0.000 (0.000)	0.005*** (0.000)
Agreeableness (Z)	-0.034*** (0.003)	-0.025*** (0.004)	-0.007*** (0.000)	-0.013*** (0.000)
Neuroticism (Z)	0.066*** (0.008)	0.038*** (0.011)	0.012*** (0.000)	0.017*** (0.001)
Sex $\times$ Ethnicity FE	Yes	Yes	Yes	Yes
Age + Age ²	Yes	Yes	Yes	Yes
R ²	0.006	0.010	0.102	0.118
Observations	777,747	161,825	777,747	161,825
Top20-Bottom20	0.021	0.016	0.006	0.008
AUC	0.575	0.633		
E(Y)	0.079		0.038	

Table C5: ROC Comparison for Thick versus Thin Files

Entries report the difference in true-positive rate (TPR) between the *Features+Vantage* model and the *Vantage-only* model at a fixed false-positive rate (FPR): $\Delta\text{TPR} = \text{TPR}_{\text{Features+Vantage}} - \text{TPR}_{\text{Vantage}}$. Positive values mean *Features+Vantage* attains higher sensitivity at the same FPR (local ROC dominance). Predictions are from logit models estimated on the same sample: (i) Vantage score (baseline) and (ii) facial delinquency score plus Vantage. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences an increase in the number of tradelines that are 30-days-or-more past-due in the 12 months immediately preceding t (the window $[t - 12, t]$). *Thin files* are borrowers with one or fewer trade lines as of $t - 2$; *Thick files* have two or more. Differences are estimated via binormal ROC regression with 100 bootstrap repetitions; standard errors in parentheses. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

	Full sample (1)	Thin files (2)	Thick files (3)
FPR = 0.10	0.040*** (0.001)	0.121*** (0.004)	-0.002 (0.002)
FPR = 0.30	0.035*** (0.001)	0.129*** (0.004)	0.004*** (0.001)
FPR = 0.50	0.018*** (0.001)	0.062*** (0.003)	0.006*** (0.001)
FPR = 0.70	0.005*** (0.001)	0.007*** (0.002)	0.005*** (0.001)
FPR = 0.90	-0.001** (0.000)	-0.007*** (0.001)	0.002*** (0.000)
Observations	971151	167154	803997

Figure C1: Number of Tradelines

This figure shows model performance across the borrowers' credit lifecycle, proxied by the number of tradelines on their credit report as of $t - 2$. It reports the Area Under Curve (AUC) for three prediction models: facial features only (red circles), Vantage score only (green squares), and facial features combined with Vantage (yellow diamonds), all estimated with demographic and age controls. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences an increase in the number of tradelines that are 30-days-or-more past-due in the 12 months immediately preceding t (the window $[t - 12, t]$).

11

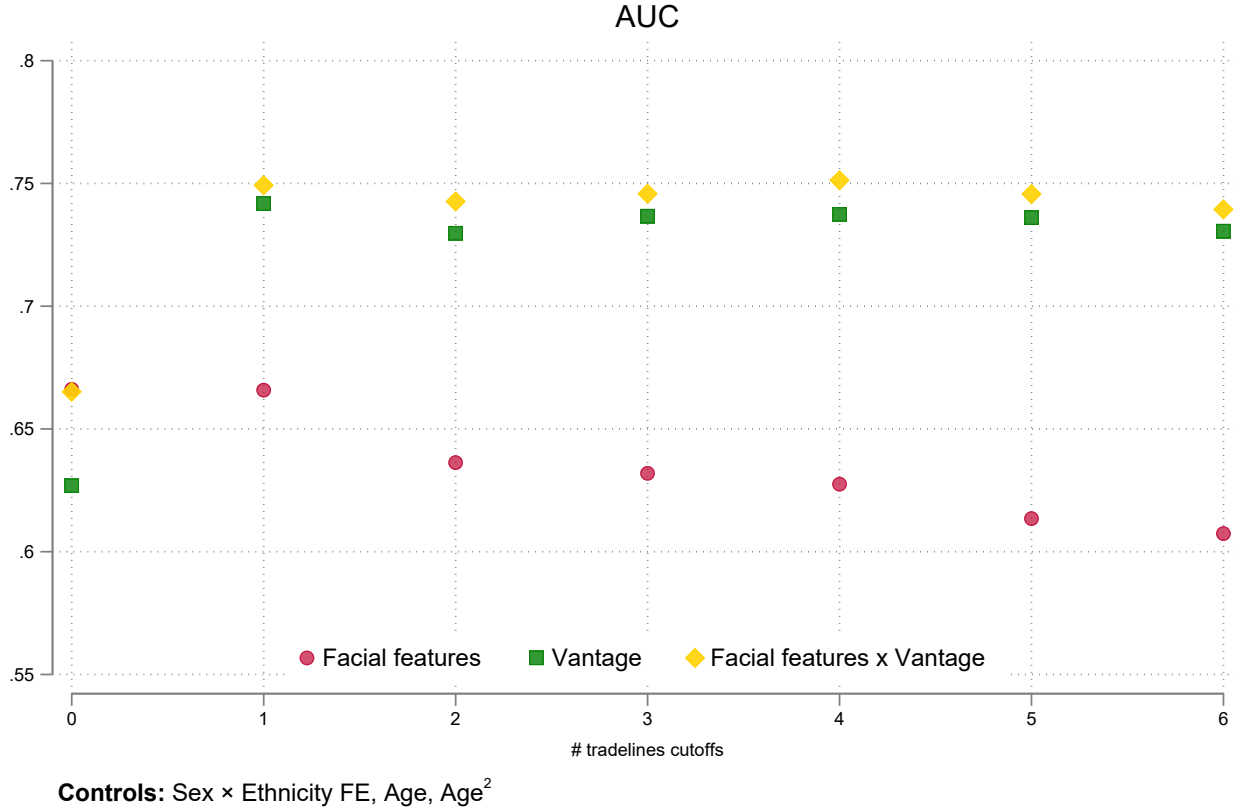


Figure C2: Borrower Age

This figure shows model performance across the borrowers' credit lifecycle, proxied by their age as of $t - 2$. It reports the Area Under Curve (AUC) for three prediction models: facial features only (red circles), Vantage score only (green squares), and facial features combined with Vantage (yellow diamonds), all estimated with demographic and age controls. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences an increase in the number of tradelines that are 30-days-or-more past-due in the 12 months immediately preceding t (the window $[t - 12, t]$).

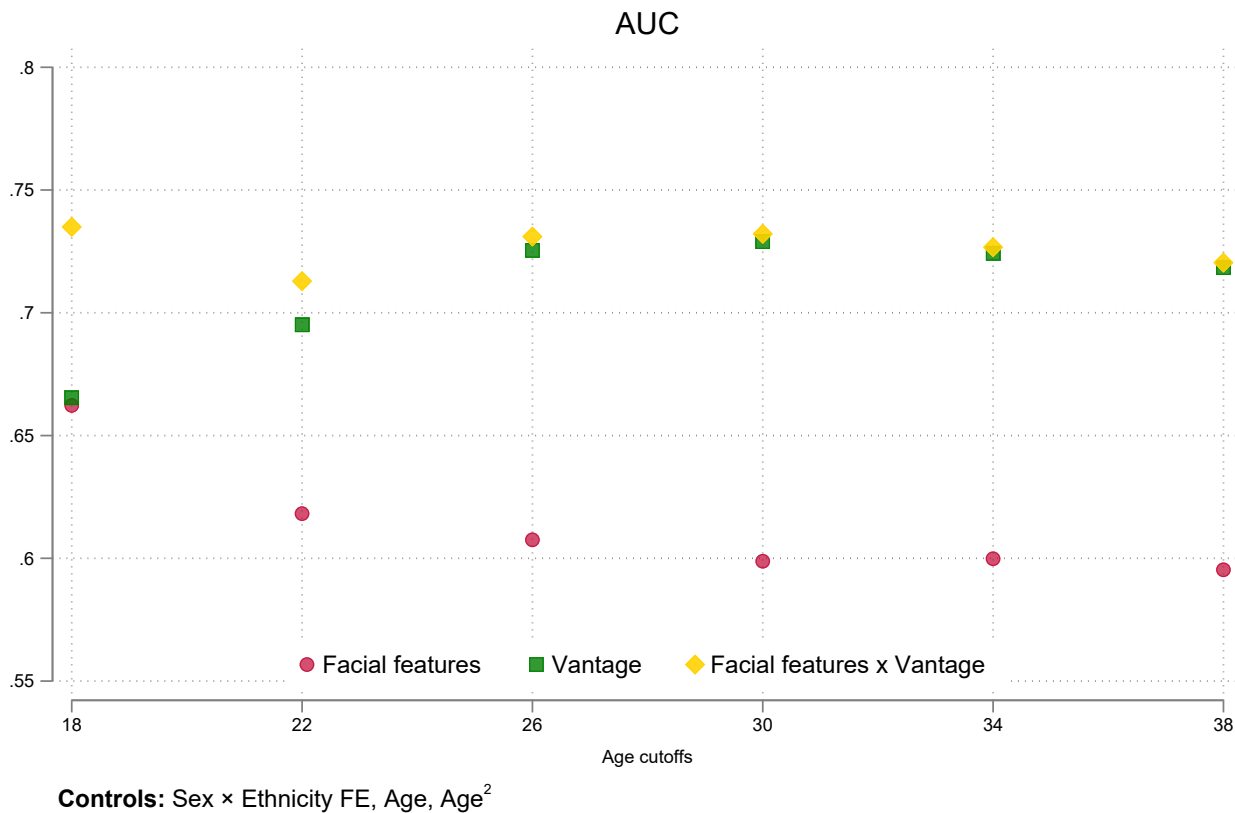


Figure C3: Borrower Income

This figure shows model performance across the borrowers' credit lifecycle, proxied by their income as of $t - 2$. It reports the Area Under Curve (AUC) for three prediction models: facial features only (red circles), Vantage score only (green squares), and facial features combined with Vantage (yellow diamonds), all estimated with demographic and age controls. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences an increase in the number of tradelines that are 30-days-or-more past-due in the 12 months immediately preceding t (the window $[t - 12, t]$).

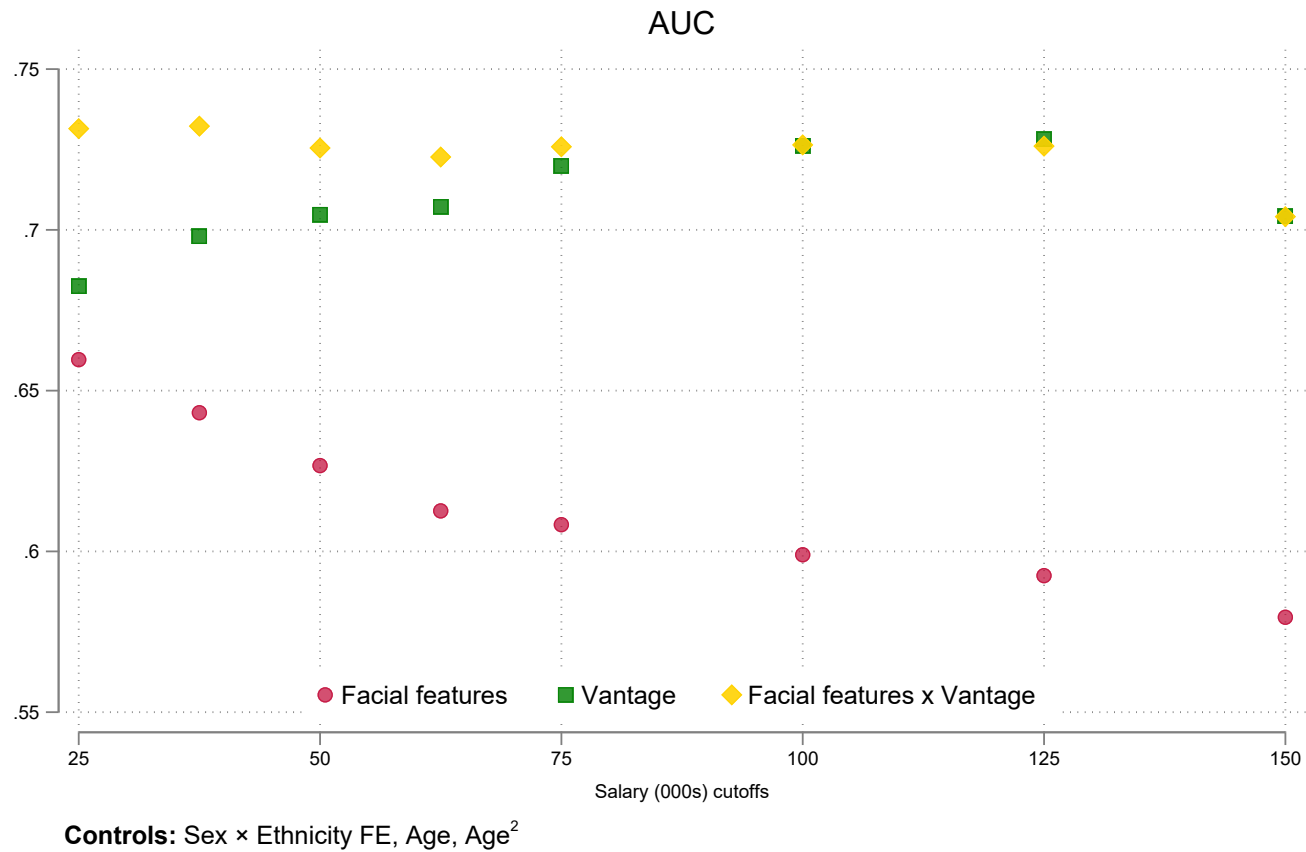


Figure C4: ROC Comparison

This plot shows Receiver Operating Characteristic (ROC) curve to measure the performance of the Vantage (blue) versus the Vantage combined with facial features (red). The ROC shows the trade-off between the true-positive rate (sensitivity) and the false-positive rate (1-specificity) at various decision thresholds of the two models. The ROC is always above the reference 45 degree line, with a more top-left curve indicating better performance in distinguishing true positives and true negatives. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences an increase in the number of tradelines that are 30-days-or-more past-due in the 12 months immediately preceding t (the window $[t - 12, t]$).

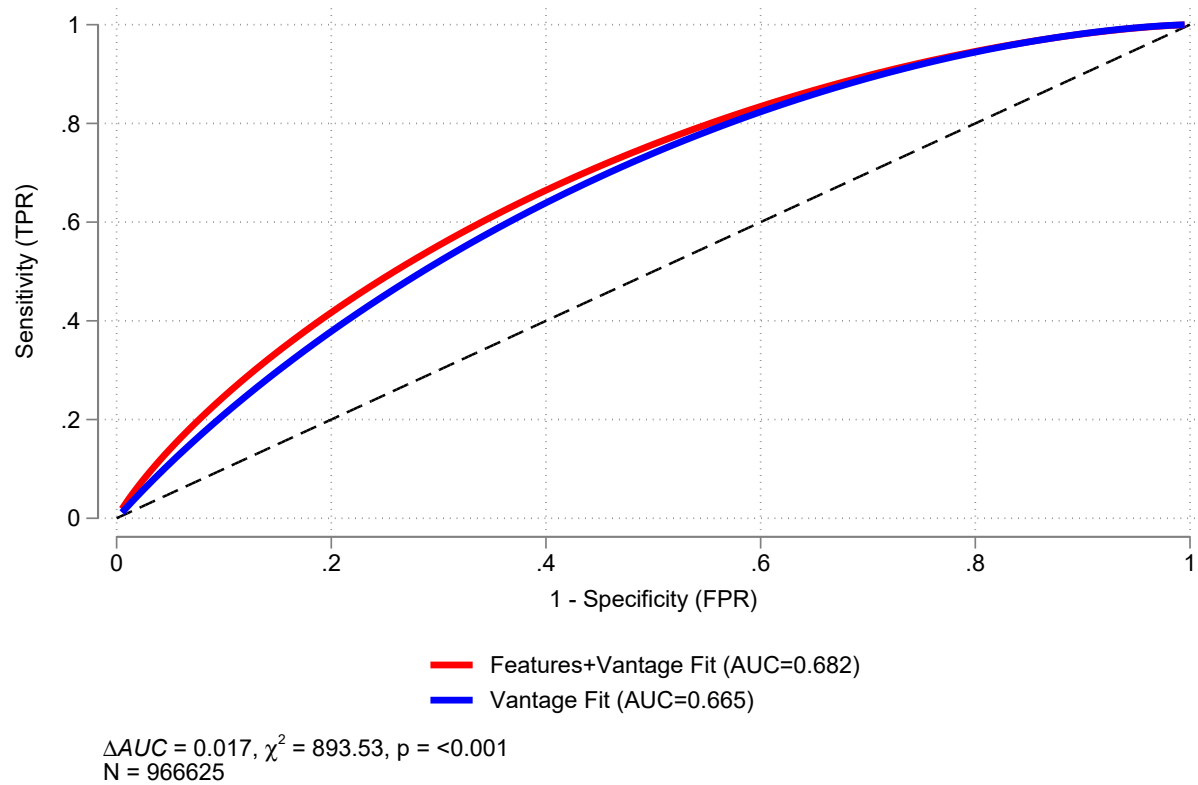
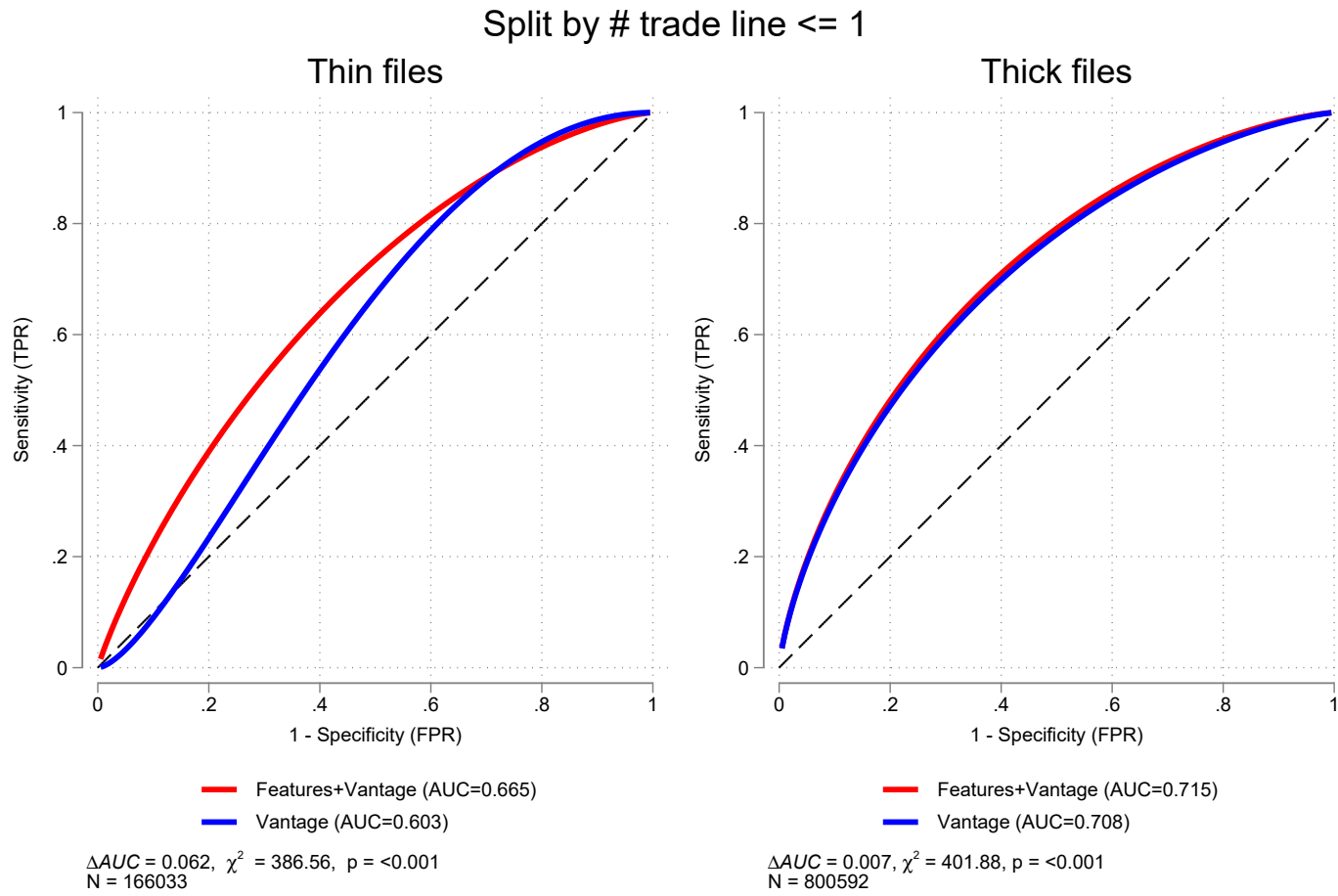


Figure C5: ROC Comparison for Thick versus Thin Files

This plot shows Receiver Operating Characteristic (ROC) curve to measure the performance of the Vantage (blue) versus the Vantage combined with facial features (red). The ROC shows the trade-off between the true-positive rate (sensitivity) and the false-positive rate (1-specificity) at various decision thresholds of the two models. The ROC is always above the reference 45 degree line, with a more top-left curve indicating better performance in distinguishing true positives and true negatives. We plot the ROC separately for thin versus thick files. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences an increase in the number of tradelines that are 30-days-or-more past-due in the 12 months immediately preceding t (the window $[t - 12, t]$).



D. ROBUSTNESS: ANALYSIS FOR A PREDOMINANT BORROWER GROUP

Table D1: Vantage Score Prediction

This table regresses Vantage scores on the facial credit score, log salary and U.S. undergraduate school rank. Variables denoted with (Z) are standardized to have mean 0 and standard deviation 1. Age controls are included where indicated. *Top20-Bottom20* is the difference between the average predicted Vantage score of the top and bottom quintile of individuals based on the facial credit score, expressed in standard deviation units. Robust standard errors are reported in parentheses. Significance levels are indicated by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

	Vantage score (Z)				
	(1)	(2)	(3)	(4)	(5)
Facial credit score (Z)		0.289*** (0.002)			0.128*** (0.002)
Log salary (Z)			0.292*** (0.002)		0.135*** (0.002)
US school rank (Z)				-0.110*** (0.002)	-0.071*** (0.002)
No US school rank				-0.049*** (0.004)	-0.093*** (0.003)
Age + Age ²	Yes	No	No	No	Yes
R2	0.130	0.083	0.087	0.009	0.167
Observations	382,298	382,298	382,298	382,298	382,298
Top20-Bottom20		0.794			0.353

Table D2: Delinquencies Prediction

This table regresses Delinquencies on the facial delinquencies score, log salary, U.S. undergraduate school rank, and lagged Vantage score. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30-days-or-more past-due episode on a tradeline at some point in the 12 months immediately preceding t (the window $[t - 12, t]$), conditional on being current at $t - 24$. Variables denoted with (Z) are standardized to have mean 0 and standard deviation 1. Age controls are included where indicated. *Top20-Bottom20* is the difference between the average predicted delinquencies score of the top and bottom quintile of individuals based on the facial delinquencies score. Robust standard errors are reported in parentheses. Significance levels are indicated by $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

	Delinquency						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Facial delinquency (Z)		0.017*** (0.000)				0.016*** (0.000)	0.009*** (0.000)
Log salary (Z)			-0.001*** (0.000)			0.011*** (0.000)	0.001** (0.000)
US school rank (Z)				0.009*** (0.000)		0.005*** (0.000)	0.004*** (0.000)
No US school rank				0.005*** (0.001)		0.004*** (0.001)	-0.001 (0.001)
Vantage score (Z, t-2)					-0.036*** (0.000)	-0.039*** (0.000)	-0.053*** (0.001)
Age + Age ²	Yes	No	No	No	No	No	Yes
R2	0.002	0.005	0.000	0.001	0.020	0.027	0.038
Observations	319,187	319,187	319,187	319,187	319,187	319,187	319,187
Top20-Bottom20		0.045				0.043	0.025
AUC	0.557	0.580	0.511	0.541	0.676	0.694	0.719

Table D3: Thick versus Thin Files

This table splits the sample by credit file thickness. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30-days-or-more past-due episode on a tradeline at some point in the 12 months immediately preceding t (the window $[t - 12, t]$), conditional on being current at $t - 24$. We regress *Delinquency* on the facial delinquencies score and lagged Vantage score. A *thick file* is defined as having two or more credit lines, and a *thin file* has one or zero trade lines as of $t - 2$. All specifications include age controls where indicated. Variables denoted with (Z) are standardized to have mean 0 and standard deviation 1. *Top20-Bottom20* is the difference between the average predicted delinquencies score of the top and bottom quintiles based on the corresponding facial score within each subsample. Robust standard errors are reported in parentheses. Significance levels are indicated by $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

	Delinquency					
	Thick file			Thin file		
	(1)	(2)	(3)	(4)	(5)	(6)
Facial delinquency (Z)	0.017*** (0.000)		0.009*** (0.000)	0.010*** (0.001)		0.008*** (0.001)
Vantage score (Z, t-2)		-0.073*** (0.001)	-0.071*** (0.001)		-0.013*** (0.001)	-0.012*** (0.001)
Age + Age ²	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.005	0.050	0.051	0.008	0.010	0.012
Observations	270,386	270,386	270,386	56,577	56,577	56,577
Top20-Bottom20	0.046		0.023	0.027		0.023
AUC	0.572	0.741	0.742	0.636	0.644	0.669
E(Y)		0.063			0.029	

Figure D1: ROC Comparison

This plot shows Receiver Operating Characteristic (ROC) curve to measure the performance of the Vantage (blue) versus the Vantage combined with facial features (red). The ROC shows the trade-off between the true-positive rate (sensitivity) and the false-positive rate (1-specificity) at various decision thresholds of the two models. The ROC is always above the reference 45 degree line, with a more top-left curve indicating better performance in distinguishing true positives and true negatives. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30-days-or-more past-due episode on a tradeline at some point in the 12 months immediately preceding t (the window $[t - 12, t]$), conditional on being current at $t - 24$.

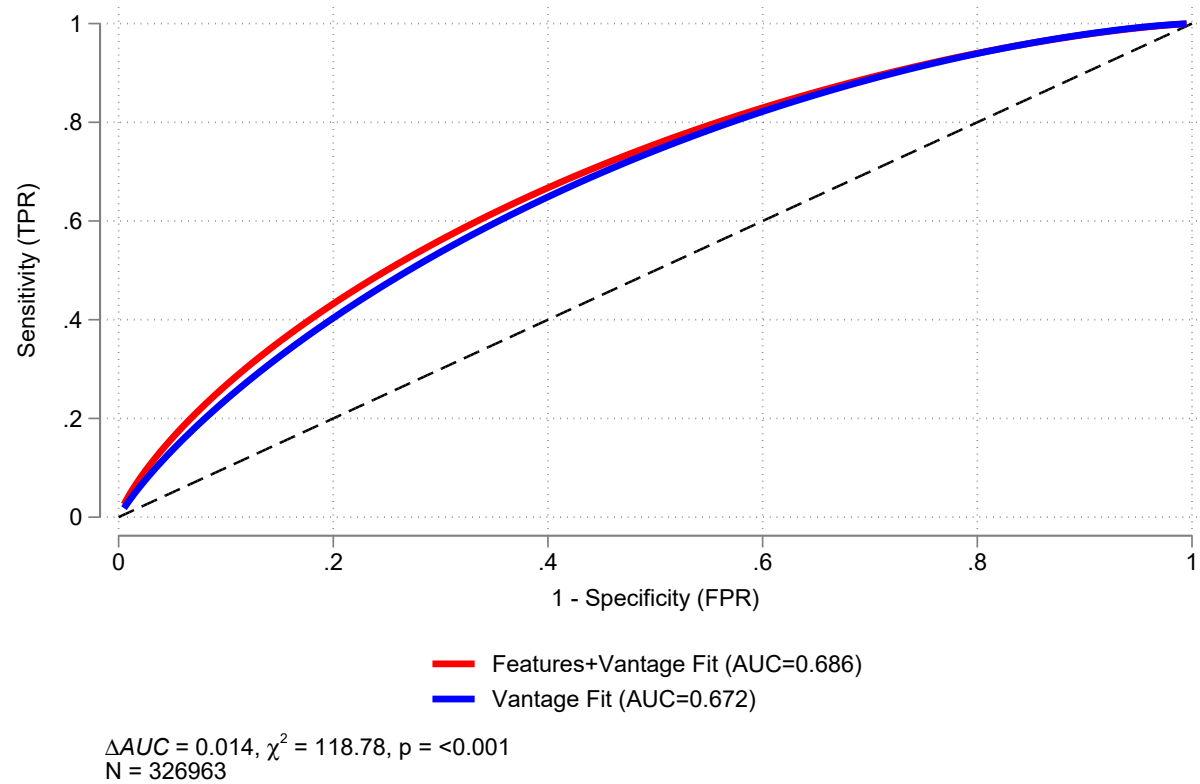


Figure D2: ROC Comparison for Thick versus Thin Files

This plot shows Receiver Operating Characteristic (ROC) curve to measure the performance of the Vantage (blue) versus the Vantage combined with facial features (red). The ROC shows the trade-off between the true-positive rate (sensitivity) and the false-positive rate (1-specificity) at various decision thresholds of the two models. The ROC is always above the reference 45 degree line, with a more top-left curve indicating better performance in distinguishing true positives and true negatives. We plot the ROC separately for thin versus thick files. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30-days-or-more past-due episode on a tradeline at some point in the 12 months immediately preceding t (the window $[t - 12, t]$), conditional on being current at $t - 24$.

