Facing Default?*

Pranav Garg
Yale

Marius Guenzel
Wharton

Shimon Kogan
Wharton/Reichman

Marina Niessner
Indiana

Kelly Shue
Yale

March 26, 2026

Preliminary. Do not circulate without permission.

Abstract

Soft information, such as personality traits and economic preferences, is likely to shape borrowing and repayment behavior, yet remains absent from the administrative data that underpin credit decisions. We assemble a novel dataset linking LinkedIn profile images to Experian credit records for over one million U.S. borrowers. Facial embeddings extracted using machine learning (ML) predict delinquency risk with meaningful incremental power beyond credit scores, demographics, income, and education. This gain is largest among younger, lower-income, and thin-file borrowers, groups for which traditional scoring is least informative. We then trace one underlying mechanism to personality and economic preferences. Using a survey of over 8,500 individuals, we train an ML model linking facial images to validated Big 5 and preference measures, and show that these measures, when applied to our LinkedIn–Experian sample, predict both delinquency and credit scores. Together, these findings highlight a tension between expanding credit access for underserved borrowers and safeguarding privacy and individual autonomy.

Keywords: Credit risk, soft information, AI and machine learning, personality, economic preferences

*This draft has benefited from comments by Tony Cookson as well as presentations at the 2025 Boulder Summer Conference on Consumer Financial Decision Making, the University of Maryland, the FRA Conference (early ideas session). 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. The survey used to develop the personality trait and economic preference models was deemed exempt by the Institutional Review Board at the University of Pennsylvania under protocol number 857435. We gratefully acknowledge financial support from the Wharton AI & Analytics Initiative and the Wharton Behavioral Lab.

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. Canonical models of credit markets such as Jaffee and Russell (1976), Stiglitz and Weiss (1981) and Sharpe (1990) formalize these questions as a problem of imperfect information, where lenders must infer borrowers' underlying default risk. Testing these theories empirically requires identifying the borrower types that underpin credit market outcomes (e.g.,"honest" versus "dishonest," "high" versus "low" quality, and "safe" versus "risky"). Modern credit scoring draws on extensive "hard" information to approximate these types, but widely used models vary substantially in accuracy across borrower populations (e.g., Blattner and Nelson (2021), Bartlett et al. (2022)). Scores are most informative for borrowers with long and stable credit histories (typically older, higher-income, and thick-file consumers with extensive credit records), and are least informative for "credit invisibles" (such as younger, lower-income, and thin-file consumers).

Beyond hard observables, a meaningful share of residual 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, as well as underlying economic preferences such as risk tolerance and time preferences. These factors shape day-to-day financial behavior but are rarely captured in administrative data. Alternative behavioral credit data, such as digital footprints, social networks, and transaction patterns, are increasingly used to proxy for these dimensions (e.g., Lin et al. (2013), The Economist (2016), 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 serve as a scalable proxy for the "soft" attributes that traditional credit models overlook. We assemble a large and novel dataset merging 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 these images and evaluate their predictive power for both realized delinquency and ex-ante credit risk, as measured by credit scores. All performance is evaluated strictly out of sample

and benchmarked against a rich set of demographic, employment, and education controls. To shed light on underlying mechanisms, we further use a separate dataset of facial information and survey responses to train ML models that generate predictions of personality traits and economic preferences, such as risk tolerance and time preferences. We apply the models' predictions to the main data to examine their relationship with credit outcomes. Our objective is not to establish causal relationships or to advocate the use of facial data in credit decisions; rather, we document predictive relationships to better understand where residual information about credit risk resides (Ludwig and Mullainathan (2024)).

Facial features may be informative about borrowers through several non-mutually exclusive channels. As we discuss in more detail below, a growing body of research in genetics, psychology, and machine learning has documented that facial appearance contains information related to personality traits through genetic, developmental, and behavioral pathways (see, e.g., Claes et al. (2014), Vukasović and Bratko (2015), Kachur et al. (2020), Guenzel et al. (2025)). To the extent that these traits and economic preferences shape how individuals make financial decisions, facial features may capture variation in underlying credit risk that is not reflected in traditional data. Importantly, the facially inferred measures we use are trained to predict self-reported personality rather than others' perceptions of personality based on appearance.[1] While our baseline results capture the combined predictive signal, including in comparisons within demographic groups, we provide evidence that personality traits and economic preferences account for a meaningful portion of this signal.

Our academic evaluation of these tools is grounded in emerging real-world practices. In many settings, lenders have already begun incorporating facial technologies into credit processes. For example, roughly 70% of Brazilian banks now use facial recognition for client identification (e.g., Agência Brasil, 2024) and Pakistan's central bank explicitly endorses artificial intelligence (AI)-based facial recognition for digital onboarding (e.g., State Bank of Pakistan, 2025). In China, major institutions have piloted face-based analysis during loan interviews to detect misrepresentation and fraud risk (e.g., Financial Times, 2018, Biometric

[1]There is also related work showing that perceptions of personality based on facial appearance can reflect superficial visual cues rather than underlying traits (Todorov et al. (2005), Willis and Todorov (2006)). While self-reported personality and such perceptions are not fully orthogonal, the correlation is typically weak when assessed by strangers (Connolly et al. (2007)).

Update, 2018, Ping An, 2019). More broadly, the use of alternative data is expanding rapidly, including in the United States, where non-bank digital lenders now originate a substantial share of personal loans. While these lenders rely on nontraditional data sources, they generally do not yet incorporate visual information. Given these industry trends, independent academic evaluation is both timely and necessary.

To evaluate the role of soft information conveyed through facial features, we construct a large-scale and novel dataset combining LinkedIn profile images and employment histories from Revelio Labs, with validated demographics from voter files, and credit bureau records from Experian. The merged dataset covers more than one million U.S. borrowers. The analysis focuses on the 2019 credit bureau vintage, augmented with earlier snapshots to construct lagged covariates. The resulting data combine rich information on credit outcomes, including Vantage scores and delinquency, with detailed demographic, education, and labor market characteristics, allowing us to study the predictive content of facial features in a large-scale, real-world setting.[2]

To operationalize the ML predictions, we extract facial embeddings from profile images using a pre-trained deep learning model designed to produce stable, person-level representations (Deng et al., 2019), and use these embeddings to train ML models that predict credit outcomes under strict out-of-sample evaluation. We train models based solely on facial features and models that combine facial embeddings with traditional credit information, allowing for flexible, nonlinear interactions between "soft" and "hard" signals. This approach enables us to isolate the predictive content of facial features and assess their incremental value beyond standard measures.

To shed light on underlying mechanisms, we further train separate ML models using a newly collected survey dataset that combines facial images extracted from short video recordings with validated questionnaires on personality (Big Five) and economic preferences (risk tolerance and patience), from over 10,000 participants recruited on Prolific and covering a wide range of ethnicities, ages, and backgrounds. We then apply these models to generate trait and preference measures for borrowers in our credit sample and assess how much of the

[2]A key strength of the data is their scale and granularity. While the sample is restricted to registered voters and LinkedIn users, and thus tilts toward a more engaged and higher-quality borrower population, it remains broadly representative along key demographic dimensions such as sex, race, and ethnicity.

predictive signal these traits and preferences account for.

Our results can be summarized in five main findings. First, facial-feature-based predictions contain economically meaningful information about credit risk beyond standard observables. In the full sample, a one standard deviation increase in our ML-derived facial delinquency score is associated with a 2.6 percentage point (pp) increase in realized delinquency (relative to a 7% mean). This relationship remains robust after controlling for demographics as well as the Revelio-based human capital measures such as income and education, which themselves are novel delinquency predictors but are associated with substantially smaller effects. To further quantify the magnitude, borrowers in the top quintile of the facial-feature-based predicted distribution are 7.1 pp more likely to become delinquent than those in the bottom quintile (3.7 pp after controls). In terms of classification performance, the facial-feature-only model achieves an AUC of 0.607, meaning it correctly ranks a randomly selected defaulter above a non-defaulter about 61% of the time. This exceeds the predictive performance of models based on demographics or human capital variables, which achieve lower AUCs in the 0.52–0.59 range.

Second, we show that facial information complements traditional credit scores. Lagged Vantage scores alone provide substantial predictive power, achieving an AUC of 0.677 and reflecting the strength of traditional "hard" information. Augmenting Vantage scores with facial features further improves performance, increasing AUC to 0.692, while a flexible model that jointly incorporates facial embeddings and Vantage scores achieves an AUC of 0.723, outperforming both standalone and additively combined specifications. This suggests that nonlinear interactions between "soft" and "hard" information are important. Consistent with this, ROC curves show that the combined ML model dominates the Vantage only model, delivering higher true positive rates for any given false positive rate.

Third, we show that the incremental value of facial features is concentrated in informationally sparse environments. For borrowers with thin credit files, adding facial information increases AUC from 0.673 to 0.696 (2.3 pp increase), and by 5.5 pp in the combined ML model, whereas gains for thick-file borrowers are close to negligible. Similar patterns emerge across age and income sorts: facial features improve AUC by 5–9 pp for the youngest and lowest-income borrowers, with effects declining substantially for older, higher-income indi-

viduals. This pattern is also reflected in ROC curves: for thin-file borrowers, the combined model's ROC curve lies substantially above that of the Vantage-only model, particularly at low false positive rates, indicating stronger discrimination in the region where false positives are kept low. In contrast, for thick-file borrowers, the curves are nearly indistinguishable.

Fourth, facial features also predict ex-ante measures of credit risk. Facial-based credit scores explain 12–13% of the variation in Vantage scores, exceeding the explanatory power of income and education, and increase overall fit to nearly 20% when combined with standard controls. This indicates that facial embeddings capture information that aligns with how credit risk is assessed in practice, while the overlap between the two is modest.

Importantly, none of the above patterns are driven by cross-group demographic differences. When we replicate the analysis within a predominant demographically homogeneous subsample, eliminating variation across sex and race, the results closely mirror those in the full sample. Facial-feature-based predictions retain similar predictive power for both delinquency and credit scores, and their incremental value remains concentrated among borrowers with sparse credit histories.

Finally, we provide evidence on underlying mechanisms. We show that personality traits and economic preferences, inferred from facial images, predict both realized delinquency and credit scores with economically meaningful magnitudes. Borrowers in the top quintile of personality-based predicted delinquency are 2.8 pp more likely to become delinquent than those in the bottom quintile, even after controlling for standard observables. Economic preferences display similarly intuitive patterns: higher predicted risk-taking is associated with higher delinquency, while greater predicted patience is associated with lower delinquency. Consistent with earlier results, the predictive power of these measures is strongest among thin-file borrowers. Taken together, these findings indicate that personality traits and economic preferences represent one channel through which facial embeddings encode credit-relevant "soft" information.

We reiterate that our goal is to evaluate the predictive power of facial embeddings in credit markets, where similar tools could be widely adopted. While such tools could, in principle, reduce the regressive effects of traditional risk models, we do not advocate their use in credit screening, as inferring personality or risk from facial features constitutes a

form of statistical discrimination (Phelps (1972)). Although algorithms can be programmed to equalize score distributions across race, ethnicity, or gender, doing so involves tradeoffs across competing fairness criteria. Moreover, an even 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.

Related Literature. Our paper contributes to four strands of literature in consumer credit, household finance, and the economics of information, as well as the growing role of big data in credit markets. First, we contribute to work on information and screening in credit markets. A large literature emphasizes the role of both hard and soft information in lending decisions (see, e.g., Petersen and Rajan (1994, 2002), Liberti and Petersen (2019)). More recently, advances in data availability have enabled the use of alternative data sources to complement traditional credit scores. Berg et al. (2020) show that simple digital footprint variables match or exceed credit bureau scores in predicting default, and Berg et al. (2022) survey this broader fintech lending landscape. We extend this literature by introducing facial embeddings as a novel source of borrower information and showing their incremental predictive power over traditional scoring technology.

Second, we contribute to work linking personality traits and economic preferences to financial outcomes. Ru and Schoar (2016) show that lenders attempt to screen for traits such as patience and self-control through indirect mechanisms. Einav et al. (2016) distinguish between contract-invariant characteristics and behavioral responses embedded in observed risk scores. We build on this distinction by focusing on the contract-invariant component of heterogeneity. Using tools from computer vision, we map facial features into validated measures of personality and preferences, providing a scalable proxy for stable individual characteristics that are otherwise difficult to observe. Consistent with this interpretation, Indarte et al. (2026) provide structural evidence that behavioral preferences drive heterogeneity in

consumption and repayment decisions.

Third, we engage with the literature on algorithmic screening and discrimination in credit markets. Fuster et al. (2022) show that machine learning models can widen disparities by indirectly encoding protected characteristics, while Bartlett et al. (2022) find that algorithmic lenders can reduce disparate impact relative to face-to-face channels. Greenwald et al. (2024) highlight challenges in measuring discrimination using demographic proxies. More broadly, this literature emphasizes inherent trade-offs across competing fairness criteria (Kleinberg et al. (2016), Chouldechova (2017)). These concerns are central in our setting, where facial features may proxy for demographics rather than underlying traits. We address this by examining within-group variation, while highlighting broader privacy and ethical considerations.

Fourth, a growing literature in genetics, psychology, machine learning, and economics documents that facial features are informative about personality traits through genetic and developmental pathways (e.g., Claes et al. (2014), Vukasović and Bratko (2015), Kachur et al. (2020), Peterson et al. (2022), Athey et al. (2022), Guenzel et al. (2025)). Separately, a large literature in psychology and economics links Big Five personality traits and economic preferences to financial behavior, including savings, borrowing, and default (e.g., Heckman et al. (2006), Roberts et al. (2007), Donnelly et al. (2012), Brown and Taylor (2014)). Structural models of household finance further interpret these preferences as key determinants of intertemporal choices and repayment behavior (e.g., Chatterjee et al. (2007), Livshits et al. (2007), Meier and Sprenger (2010)). We bridge these three strands by showing that facial embeddings capture variation in personality traits and economic preferences that is predictive of credit outcomes, providing a scalable proxy for latent determinants of financial behavior.

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 Big 5 and economic preference measures. Section 3 outlines our estimation strategy, cross-validation procedures, and specifications. Sections 4 and 5 present the main results on delinquency prediction, heterogeneity analyses, and underlying mechanisms related to personality and economic preferences. Section 6 concludes.

2. Data and ML Methodology

We assemble an individual-level dataset by linking professional profiles from Revelio Labs (which include LinkedIn photographs and employment histories), validated demographics from voter registration files (L2), and credit bureau records from 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.[3] The final sample contains approximately 1.1 million individuals.

2.1 Revelio Labs Data

Our LinkedIn dataset comes from Revelio Labs, a commercial 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' address histories and exact dates of birth – both required to link to credit-bureau records – we match the Revelio/LinkedIn data with L2 voter files, a continuously updated national voter-registration database compiled from state records covering more than 200 million registered voters. L2 provides a persistent person identifier and standardized demographic fields. Linking to L2 restricts our sample to individuals who have registered to vote in the United States.

We merge Revelio/LinkedIn to L2 in three stages. First, we prepare the LinkedIn records for matching. We standardize names and construct a person-by-time panel of state-level locations from each individual's education and work histories. We estimate year of birth from college start date; individuals without a reported start year are dropped, as approximate age is required for matching. Together, these steps produce the identifying fields (name, state-date residence history, and approximate age) needed to link to L2.

Second, we match each LinkedIn profile to L2 on full name, state-by-date residence history, and approximate age (± 2 years). Because individuals may reside in multiple states over the sample period, we split each person's residence history into state-specific spells

[3] Experian pulls data on the final Wednesday or Saturday of each month. Given the annual nature of our data, we obtain snapshots as of the final Wednesday or Saturday in December 2017 and December 2019.

and match each spell separately. Matches are based on fuzzy name similarity and require overlapping state-date ranges, allowing a ±365-day window to account for lags in updating LinkedIn profiles.[4] This step yields 3,540,960 unique matched individuals.

Third, we apply a large-language-model-based validation to the intermediate matched file, reclassifying 2,181,578 records as true positives and discarding the remainder as false matches.[5] Ambiguous pairs are dropped rather than risking false positives. The resulting 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). Appendix A provides further details on all data-cleaning and merging steps.

2.2 Experian Credit Data

We transmit standardized identifiers (first, middle, last, suffix, date of birth, most recent address) for all 2,181,578 validated Revelio-L2 matches to Experian and request annual credit snapshots. Experian successfully matched 1,757,090 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, pre-computed summary fields that flag whether a borrower has experienced a past-due event within rolling lookback windows of 6, 12, 24, and 36 months. We then restrict to individuals with a LinkedIn profile photo (1,501,894), and usable facial predictions from both our credit-outcome and personality models, yielding a final analysis sample of 1,097,588 individuals.

Age is obtained from L2; sex and ethnicity are provided by Revelio, along with human-capital proxies: imputed salary, an indicator for missing school rank, and a numeric school-rank measure where higher values indicate worse prestige. These are computed as of 2017 or the earliest available data prior to 2017. Credit variables include the contemporaneous VantageScore, utilization rate, tradeline counts, and delinquency indicators.

[4]For example, a 47-year-old individual who resided in Nebraska from September 2017 to May 2019 and then in Connecticut from June 2019 onward would generate two blocks: {Doe, John, 47, Nebraska, May 2017–June 2019} and {Doe, John, 47, Connecticut, June 2019–Present}. The Nebraska block would match to an L2 record {Doe, John, 46, Nebraska, Jan 2017–Dec 2019} because the names are similar, ages are within ±2 years, and the state-date ranges overlap within the ±365-day tolerance.

[5]The LLM accommodates abridged middle names and tolerates spelling errors up to one per every seven letters in a name.

Our primary delinquency outcome equals one if the borrower experiences any 30-days-or-more past-due episode on any tradeline during the twelve months preceding the 2019 vintage date, conditional on being current on all tradelines twenty-four months prior (i.e., in 2017, referred to as "t-2" throughout). This definition captures "first-time" delinquencies: a switch from current to a 30+ days-past-due (DPD), rather than "serious delinquency" typically associated with being 90+ days-past-due.[6] We exclude individuals who were delinquent at $t-2$, so the control group consists of individuals who are current in both periods. As a robustness measure, we code an indicator for whether the number of tradelines that are 30+ days-past-due increases over the same twelve month window. Both definitions exploit the bureau's rolling delinquency 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 tradelines in 2017 equals 0 or 1, and is otherwise classified as having a thick credit file.[7]

2.3 ML Facial Feature Extraction and Predictions

We construct facial-feature-based predictors of credit outcomes in two steps. First, we map each LinkedIn photo into a fixed-length facial embedding using a pretrained deep convolutional neural network. Second, we feed these embeddings into a separate supervised machine learning model to predict credit outcomes with strict out-of-sample evaluation.

In the embedding step, we process each LinkedIn profile photograph through a deep convolutional neural network pretrained on approximately 12 million face images spanning 600 thousand unique identities, optimized for face recognition.[8] The network maps each face into a 512-dimensional embedding vector. Crucially, this embedding is extracted from a general-purpose face recognition model and is not trained on any credit outcome, so the

[6] Results, available upon request, remain robust to using this alternate 90+ DPD definition.

[7] Results, available upon request, remain robust to defining credit file thinness as 2 or fewer, instead of 1 or fewer.

[8] Our pipeline uses a 50-layer residual network (ResNet-50) fine-tuned with InsightFace's ArcFace loss function, which improves discrimination by operating in angular rather than Euclidean space (e.g., He et al. (2016), Deng et al. (2019)). The resulting 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 across our large sample.

representation is independent of our prediction targets.

In the supervised stage, we predict the borrower's contemporaneous Vantage score (continuous) and the two delinquency indicators defined earlier (binary) from the facial-feature embeddings. We consider two feature sets. In the *faces-only* specification, the model takes only the 512-dimensional embeddings as inputs. In the *faces + Vantage* specification, we augment the embeddings with the borrower's 2017 Vantage score to allow for nonlinear interactions between traditional "hard" information and visually-inferred "soft" information within a combined ML model. The motivation for this combined specification is that non-traditional signals and digital footprints can complement bureau scores in predicting default and credit outcomes (see, e.g., Duarte et al. (2012), Ravina (2020), Berg et al. (2020), Fuster et al. (2022)). A joint ML model can capture nonlinear interactions between these two information sources that additive linear specifications would miss.

Our primary supervised model uses gradient-boosted decision trees implemented via LightGBM, a method widely recognized for strong performance on structured, high-dimensional data (Ke et al. (2017)). For binary delinquency prediction, we target the area under the receiver operating characteristic curve (ROC-AUC) as our main performance metric. For continuous credit score prediction, we target RMSE.

To evaluate predictive performance, we partition the data into three folds. Each fold serves once as a held-out test set, remaining untouched until final scoring, 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 additional three-fold cross-validation, using an internal validation split and early stopping to tune hyperparameters.[9] This procedure ensures that reported performance reflects genuine out-of-sample accuracy: the algorithm never observes test individuals at any stage of training or tuning.

[9]We 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.

2.4 Photo Big 5 Personalities and Economic Preferences

Research in genetics and psychology suggests that personality traits are partially reflected in facial appearance, and that machine learning models can predict self-reported personality from photographs with modest but meaningful accuracy (e.g., Ambady and Rosenthal (1992), Kachur et al. (2020), Guenzel et al. (2025)). Motivated by this evidence, we construct image-based measures of personality traits and economic preferences by training a deep learning model to predict self-reported traits and preferences from facial photographs.

Our prediction targets fall into two domains. The first is personality, measured by the Big 5 framework—the most widely used taxonomy of personality in psychology and economics—encompassing Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism (OCEAN).[10] We refer to the image-extracted Big 5 measures as the Photo Big 5. The second domain is economic preferences, specifically patience and risk-taking, which a large literature identifies as central determinants of financial behavior including borrowing, saving, and default (e.g., Falk et al. (2018)). Although personality traits and economic preferences are conceptually distinct, they are closely related in practice: both capture stable individual dispositions that shape economic decision-making (e.g., Borghans et al. (2008), Almlund et al. (2011), Jagelka (2024)).

Our training data consist of 8,694 individuals, recruited on Prolific, for whom we collect short video recordings and administer a validated 60-item Big 5 personality questionnaire alongside single-item measures of patience and risk-taking. We select a single still frame from each video and split the sample into a training pool (7,391 individuals) and a held-out test set (1,303 individuals). The prediction target for each individual is the vector of seven scores (five personality traits and two economic preferences), each normalized to $[0, 1]$. The preference measures, being based on single survey items rather than validated multi-item scales, are noisier than the Big 5 scores; this measurement asymmetry attenuates the predicted correlations and downstream regression coefficients for patience and risk-taking.

The pipeline proceeds in two stages. In the first stage, we extract a 512-dimensional facial embedding from each image using InceptionResNetV1, a deep neural network pretrained on

[10] See, for example, Almlund et al. (2011) and Heckman and Kautz (2012) for reviews of the Big 5 in economics.

VGGFace2 for face recognition.[11] As with the ArcFace embeddings used for credit prediction, this step converts raw photographs into a structured numerical representation, but the two networks are distinct. InceptionResNetV1 is better suited to our smaller personality training sample, as it does not require the angular-margin fine-tuning designed for large-scale identity separation.

In the second stage, we train an attention-gated regression model that maps the 512-dimensional embedding into seven scores.[12] We use five-fold cross-validation within the training pool and average predictions across the five fold-specific models to produce the final scores. All reported performance is evaluated on the held-out test set.

We then apply the trained model to the LinkedIn profile photos in our main sample to obtain predicted Photo Big 5 scores and predicted economic preferences (risk-taking and patience) for each individual. These scores summarize the self-reported traits and preferences that our model extracts from facial photographs. While correlated with their self-reported counterparts, they should not be interpreted as exact measures of latent traits or preferences. In the held-out test sample, correlations between predicted and self-reported scores range from 0.13 to 0.24 across the seven measures. These magnitudes are comparable to correlations between behavioral task measures and survey-based personality assessments, which rarely exceed 0.3 (e.g., Almlund et al. (2011)).

2.5 Summary Statistics

The combined dataset incorporates individual demographic information, human capital-proxies, credit bureau data, feature-based predictions of credit outcomes, and measures of the Photo Big 5 and economic preferences for approximately 1.1 million consumers. Table 1 reports summary statistics for the person-level 2019 cross-section.

[11] The backbone network is frozen during training: its parameters are not updated. Images are resized to 160×160 pixels and normalized before extraction. Each 512-dimensional embedding is L2-normalized to unit length.

[12] The model consists of a shared block (512 to 128 dimensions) followed by seven trait-specific heads, each equipped with a learned gating mechanism that modulates the shared representation. This architecture allows each of the seven outputs to load on a different combination of the input features, so that the model can learn trait-specific and preference-specific patterns in facial appearance. We train using Huber loss with uncertainty-weighted multi-task learning, which automatically balances traits that differ in noise and difficulty. Optimization uses AdamW with learning-rate warmup and early stopping. See Appendix B for full architectural and training details.

Relative to widely used bureau datasets, the sample skews younger, with a median age of 28, and is relatively high-income, with a median salary of $66k (Lee and van der Klaauw (2010)). Borrowers are generally in the prime range: the mean VantageScore is 713 (P50 = 729; P90 = 817), placing most of the sample within the commonly referenced prime band of 661 to 780 (Goodman (2024)). 7% experience a 30+ days-past-due (DPD) delinquency over the 12 months preceding 2019, conditional on being current in 2017, and 18% meet our thin-file definition of zero or one tradelines in 2017. These composition differences relative to the general U.S. adult population are expected given that our sample requires a populated LinkedIn profile. For context, prior work often uses ≤ 3 accounts when discussing noise and information sparsity in credit reports (e.g., Blattner and Nelson (2021)). Because the sample is concentrated above the subprime threshold, severe delinquency (90+ DPD) is uncommon, and we focus on 30+ DPD events as our primary outcome. This measure captures the onset of repayment difficulty and serves as an early indicator of more persistent credit distress (Laufer and Paciorek (2022)).

Figure 1 presents correlations between human capital measures from Revelio and traditional credit scores. Panel (a) shows a positive relationship between log salary and Vantage scores, with a slope of 48.9 points per log unit. Panel (b) demonstrates that graduates of higher-ranked schools (lower numerical values) carry higher Vantage scores, with a slope of -0.052. These patterns indicate that the Revelio-based human capital measures contain credit-relevant information, raising the question of whether additional "soft" information embedded in facial features provides incremental predictive power.

3. Empirical Approach

Having described the data construction and prediction pipeline, we now turn to the empirical framework.

3.1 Facial Features and Credit Outcomes

We test whether facial embeddings predict credit-related outcomes out-of-sample, and assess the economic meaning of that predictive content once standard proxies for human

capital are held constant. Our main analysis benchmarks facial-feature-based delinquency predictions against lagged traditional credit scores. The former is our measure of "soft" information and the latter is a widely-used measure of "hard" information.

We estimate:

$$y_i = \alpha + \alpha_{j(i)} + \beta \cdot p_i^{faces} + \mathbf{\Gamma'} \boldsymbol{X_i} + \varepsilon_i \tag{1}$$

where y_i is a realized delinquency indicator for borrower i and p_i^{faces} is the out-of-sample prediction of delinquency from facial embeddings. Our preferred definition identifies borrowers with a 30+ days past-due episode in the 12 months preceding year-end 2019 ($[t-12,\ t]$), conditional on being current as of 2017 ($[t-24]$). An alternate definition identifies borrowers experiencing an increase in 30+ days past-due tradelines between 2017 and 2019 ($[t-24,\ t]$).

$\alpha_{j(i)}$ denotes fixed effects for sex $\times$ ethnicity. X_i is a vector of additional controls, including linear and quadratic age and a set of benchmarking covariates (e.g., education, income, and lagged Vantage score). All variables marked with a "(Z)" in the tables are standardized to mean zero and unit variance to facilitate comparison across coefficients. All specifications use heteroskedasticity-robust standard errors.

The coefficient of interest is β, which captures the change in realized delinquency associated with a one-standard-deviation increase in the facial-feature-based prediction. We benchmark β against coefficients on established predictors of credit outcomes, such as the lagged Vantage score, income, and education (measured as of 2017). Following the credit-risk literature, we report the area under the receiver operating characteristic curve (AUC) as our headline measure of discriminatory power (e.g., Altman et al. (2010), Iyer et al. (2016), Vallée and Zeng (2019)). We also report R^2 and the *Top20-Bottom20* gap: the difference in mean predicted outcome between borrowers in the top and bottom quintiles of the predicted score distribution. As a calibration, the Vantage-only benchmark in our data attains an AUC of 0.677, comparable to the AUC of 0.683 reported by Berg et al. (2020) for a traditional credit scoring model.

We examine heterogeneity along two dimensions. First, we estimate equation (1) separately for "credit invisibles" (i.e., low income, younger, and thin-file borrowers) versus

borrowers with long credit histories (high income, older, and thick-file borrowers). Facial-feature-based predictions should be most informative where traditional credit scoring is least effective at classifying borrowers, that is, in informationally sparse environments. To test this, we compute AUCs within bins of income, age, and file thickness, comparing a Vantage-only benchmark against a model that augments Vantage with facial embeddings.

Second, we estimate equation (1) using the Vantage score as the dependent variable to test whether facial features generate "hardened soft information." A positive coefficient β paired with a high R^2 (relative to demographic and human capital controls alone) would indicate that facial embeddings contain information that overlaps with what formal scoring models aggregate from payment histories and account behavior. This overlap matters because it implies that the visual signals could substitute for bureau data in settings where traditional scoring is unavailable or unreliable, such as for thin-file or young borrowers.

3.2 The Role of Personality and Economic Preferences

We then test whether personality and economic preferences serve as an underlying mechanism. A large literature in psychology and economics establishes that personality traits, particularly those captured by the Big 5 taxonomy, predict financial behaviors including savings, debt accumulation, and default propensity (see, e.g., Donnelly et al. (2012), Brown and Taylor (2014), Roberts et al. (2007), Heckman et al. (2006)). 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 instability and poor stress management, factors relevant to financial decision-making under uncertainty. However, existing work does not establish whether these traits carry greater predictive power in informationally sparse environments where traditional credit scoring is least effective. Separately, economic preferences such as patience and risk-taking occupy a central role in household finance models of borrowing and default, yet their empirical relationship to credit outcomes remains largely untested at scale. Our analysis addresses both questions using photo-based measures for over one million borrowers.

If facial embeddings encode personality- and preference-related information, this would

provide a natural explanation for their predictive power in credit markets. We test this hypothesis in two steps. First, we examine whether the Photo Big 5 personality scores described in Section 2.4 explain variation in realized delinquency and credit scores, both unconditionally and after controlling for traditional credit measures. Specifically, we estimate:

$$\begin{aligned} \mathrm{y}_i = \alpha + \beta_1 \cdot \text{Agreeableness}_i + \beta_2 \cdot \text{Conscientiousness}_i + \beta_3 \cdot \text{Extraversion}_i \\ + \beta_4 \cdot \text{Neuroticism}_i + \beta_5 \cdot \text{Openness}_i + \boldsymbol{\Gamma'} \boldsymbol{X_i} + \varepsilon_i \end{aligned} \tag{2}$$

where y_i is either realized delinquency or the Vantage score, all personality variables are standardized to mean zero and unit variance, and X_i denotes a vector of controls including *sex* $\times$ *ethnicity* fixed effects, age, age squared, and the lagged Vantage score. This specification isolates the incremental explanatory power of Photo Big 5 traits beyond standard demographic characteristics and traditional credit scores. Because the five predicted traits are mutually correlated, we focus on the joint specification rather than interpreting individual coefficients in isolation.

Second, we examine the role of economic preferences (e.g., Falk et al. (2018), Jagelka (2024)). Patience and risk-taking enter directly as fundamental parameters in household finance models of borrowing and default (e.g., Chatterjee et al. (2007), Livshits et al. (2007), Meier and Sprenger (2010)): the discount factor governs the intertemporal trade-off between current consumption and future debt repayment, while risk aversion governs the desire to smooth consumption against unforeseen income shocks. We estimate:

$$\mathrm{y}_i = \alpha + \beta_1 \cdot \text{Risk-Taking}_i + \beta_2 \cdot \text{Patience}_i + \boldsymbol{\Gamma'} \boldsymbol{X_i} + \varepsilon_i \tag{3}$$

where X_i is defined as above. As with the Big 5 specification, y_i is either realized delinquency or the Vantage score, and we estimate this equation with and without traditional credit controls.

4. Predicting Delinquencies and Credit Scores

We begin by examining whether the photo-derived delinquency predictions predict realized delinquencies out of sample. We evaluate classification performance, test for incremental value over human capital controls and lagged Vantage scores, and examine heterogeneity by file thickness and related measures of credit history depth. The section proceeds by testing whether photo-derived features predict Vantage scores, providing a complementary test of whether these measures capture borrowers' latent credit risk.

4.1 Facial Features Predicting Delinquency

Baseline Results. Panel A of Table 2 reports baseline classification results predicting delinquencies with facial features. As described, *Delinquency* is defined as a 30-days-or-more past-due episode on a tradeline during the 12 months immediately preceding year-end 2019 (the window spanning 2018 and 2019), conditional on being current as of 2017.[13]

Columns (1) to (3) provide benchmark specifications. Demographic fixed effects and age controls separately attain AUCs of 0.563 and 0.549, and jointly an AUC of 0.591 and an R^2 of 0.7%. Column (4) 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 predictive power beyond standard demographic characteristics.

Moving from the bottom to the top quintile of the facial delinquency distribution is associated with a 6.1 percentage point increase in predicted realized delinquency, relative to a mean delinquency probability of 7%, reflecting substantial dispersion in predicted delinquency risk across photo-derived scores.

Panel B incorporates human capital controls to assess whether facial features proxy for traditional correlates of creditworthiness such as income and education. Columns (1) and

[13] All delinquency results that follow in this and subsequent sections are equally robust to an alternate definition of delinquency and are presented in Appendix D. 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 2019 (spanning 2018 and 2019).

(2) reproduce the key specifications from Panel A, restricting the sample to individuals with available human capital controls. Columns (3) and (4) show that log salary alone yields an AUC of 0.520 while school rank yields an AUC of 0.544, indicating that both human capital measures contain information about delinquency and load in the expected direction. In column (5) we jointly include the facial delinquency score with salary and school rank and attain an AUC of 0.611, an improvement of roughly 7 percentage points over the human-capital-only models. In column (6), combining the facial delinquency score with salary, school rank, and the full set of demographic controls and fixed effects yields an AUC of 0.629 and an R^2 of 1.5%.

In this full specification, a one-standard-deviation increase in the facial delinquency score is associated with a 2.1 percentage point higher realized delinquency, and the top–bottom quintile gap is 5.8 percentage points, indicating economically meaningful dispersion in predicted delinquency risk beyond these demographic, schooling, and labor proxies. Comparing columns (2) and (6), the coefficient on the facial delinquency score is largely unchanged by the inclusion of human capital proxies, suggesting that its predictive power operates through a different channel.

Information Beyond Traditional Credit Scores. In Table 3 we add lagged Vantage scores (from 2017) to assess whether facial features provide information beyond traditional credit scores.[14] Columns (1) and (2) of Panel A, reproduce the main specifications from columns (3) and (4) of Panel A of Table 2. In column (3) we find that lagged Vantage score alone substantially predicts delinquency, achieving an AUC of 0.677 and an R^2 of 2.3%. Augmenting with the facial delinquency score increases the AUC to 0.692 and the R^2 to 3.0% (column (4)), indicating that facial features carry meaningful incremental predictive content beyond traditional scores.

In column (5), with added demographic controls we find that a one-standard-deviation increase in facial delinquency score is associated with a 1.4 percentage point higher predicted realized delinquency. The *Top20-Bottom20* metric shows that moving from the bottom to the top quintile of the facial delinquency distribution is associated with a 3.8 percentage

[14]We use the 2017 Vantage score rather than the 2019 score to avoid reverse causality from delinquencies between 2017 and 2019 mechanically affecting the 2019 Vantage score.

point difference in realized delinquency, even after controlling for lagged Vantage scores and demographics.

In Panel B we add human capital controls. The full model (facial delinquency, salary, school rank, lagged Vantage score, and demographics) achieves an AUC of 0.724 and an R^2 of 4.5%. The top-bottom quintile gap remains at 3.7 percentage points after adding human capital proxies, indicating that the facial delinquency score captures information distinct from income, education, traditional credit scores, and demographics.

Combined ML Model Using Facial Embeddings and Vantage Scores Tables 2 and 3 treat the facial delinquency score and Vantage score as separate regressors and show that combining the two outperforms either input alone. We next evaluate a combined ML model that takes the 512-dimensional facial embeddings and the borrower's 2017 Vantage score as joint inputs, allowing the algorithm to estimate nonlinear interactions between traditional "hard" information and visually inferred "soft" information.

Table 4 reports the results. In Panel A, the combined ML model achieves an AUC of 0.723 (column (2)), exceeding both the faces-only model (0.607, Table 2 column (4)) and the lagged Vantage score alone (0.677, column (4)). The combined ML model also outperforms the linear specification in Table 3 that enters the facial score and Vantage score as additive regressors (0.692), indicating that a flexible ML framework captures interactions that linear models miss. The *Top20-Bottom20* gap is 14.8 percentage points, twice the sample mean delinquency probability of 7%. Adding demographic controls raises this gap slightly to 15.1 percentage points (column (3)).

Panel B of Table 4 adds human capital controls to test whether the combined ML model's predictive gain is absorbed by salary and education. The full specification in column (5) achieves an AUC of 0.735, indicating that the gain persists. This compares to an AUC of 0.724 in the corresponding specification in Table 3 that enters facial information and Vantage scores additively rather than through a joint ML pipeline.

ROC Comparisons. The AUC-based comparison across models measures overall predictive accuracy but does not reveal at which thresholds one model outperforms another. Receiver Operating Characteristic (ROC) curves provide this finer view by tracing the tradeoff between the true positive rate (TPR) and the false positive rate (FPR) as the classifica-

tion threshold varies. A curve more bowed toward the top-left achieves higher detection of delinquent accounts for any given rate of false positives.

Figure 2 plots the ROC curves for the combined $Facial + Vantage$ ML model (red) and the lagged Vantage score-only model (blue). The combined model's curve lies above the Vantage-only model curve across the full range of thresholds, with the separation most pronounced at low false positive rates.[15]

4.2 Predictive Performance of Facial Features for "Credit Invisibles"

4.2.1 Predictive Performance by Credit File Thickness

The preceding analysis pools all borrowers. We now ask whether the incremental value of facial features varies with the richness of a borrower's traditional credit record. To do so, we exploit variation in credit file thickness, which proxies for the amount of available credit history. Thin-file borrowers represent low-information environments in which standard credit scores are less precise, whereas thick-file borrowers have more extensive credit records. If facial features provide complementary information, their incremental predictive value should be greatest where conventional signals are sparse. Accordingly, we compare the predictive performance of traditional credit scores ("hard" information) and facial-feature-based predictions ("soft" information) across thin- and thick-file borrowers. We later examine how these relationships vary by age and income.

Table 5 splits the sample by credit file thickness, extending the full-sample analysis in Panel A of Table 3. We define a thick file as having two or more credit lines, and a thin file as having one or zero tradelines as of year $t-2$.

For the thick-file sample, the AUC for the facial-feature model is 0.610, compared with 0.741 for lagged Vantage, and the joint model achieves an AUC of 0.744. Adding facial-feature predictions yields only a marginal improvement in AUC. For thin-file borrowers, by contrast, the facial-feature model achieves an AUC almost as high as the Vantage model, and adding the facial delinquency score to the Vantage-based model increases the AUC from

[15]The small difference in AUC between the figure and the numbers reported in Tables 3 and 4 reflects different Stata commands: `rocreg` for the ROC curves and AUC in the figure, and `logit` for the table AUCs.

0.673 to 0.696 (2.3 percentage points). The incremental value of facial features is therefore concentrated in the thin-file sample, where traditional credit information is most sparse and noisy.[16]

In Table 6, we repeat this comparison using our combined ML model that jointly incorporates the 512-dimensional facial embeddings and the borrower's 2017 Vantage score as inputs. All estimations include demographic controls and correspond to columns (3) and (5) in Table 4 split by credit file thickness. The pattern is the same: the gain in predictive power relative to a model without facial information is concentrated in thin-file borrowers. The AUC gain over traditional credit scores alone is 5.5 percentage points in the thin-file sample (columns (3) vs. (4)), whereas the combined model offers virtually no improvement over traditional credit scores for the thick-file sample (columns (1) vs. (2)).

4.2.2 AUC Across Credit File Thickness Cutoffs

We next extend the thin- versus thick-file comparison by tracing AUC across the distribution of trade lines. Figure 3 plots AUC by number of tradelines for three models: facial-features-based predictions alone (green circles), lagged Vantage scores alone (blue squares), and the combined $Facial + Vantage$ ML model (red diamonds). All AUCs are estimated with demographic and age controls.

At zero 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, the predictive performance of Vantage scores rises substantially (reaching about 0.76 at 6 tradelines), while the facial-features-only model remains flat at around 0.625. The combined model remains slightly above Vantage scores throughout (around 0.755-0.760). Taken together, these results indicate that facial features provide meaningful incremental predictive power when credit history is most sparse, but their marginal contribution diminishes as traditional credit information becomes available.

[16]Relatedly, the coefficient on facial delinquency drops significantly for thick-file individuals when the Vantage score is added, suggesting that much of its explanatory power overlaps with hard information conveyed in Vantage scores (columns 1 vs. 3). For thin-file individuals, by contrast, the coefficient remains largely unchanged (columns 4 vs. 6).

ROC Comparisons. As in the full sample, we examine ROC curves to complement the AUC heterogeneity analysis by credit file thickness. Figure 4 shows the ROC results split across thin- and thick-file borrowers, allowing us to examine where along the risk spectrum the facial-information ML models improve on lagged Vantage scores.

In both panels, the combined $Facial + Vantage$ ML model (red) lies above the Vantage-only classifier (blue). The separation is much more pronounced for thin-file borrowers (left panel), particularly at low false positive rates, where the red curve rises more steeply. For thick-file borrowers (right panel), the two curves are nearly indistinguishable, with only modest differences across thresholds. To quantify this contrast: at a false positive rate of 0.10, the combined model achieves a 22.6 percentage point higher true positive rate than Vantage alone for thin-file borrowers, compared with essentially zero for thick-file borrowers (Appendix Table C1). The incremental value of facial information is thus concentrated among thin-file borrowers, and especially at conservative thresholds.

4.2.3 AUC Across Age and Income Cutoffs

We next examine whether the pattern documented for file thickness, namely, that facial features add the most value where traditional information is sparse, extends to other characteristics associated with being "credit invisible," specifically young age and low income.

Figure 5 shows model performance by borrower age. For the youngest borrowers (age 18), the combined model substantially outperforms Vantage scores alone by roughly 8 to 9 percentage points in AUC, and the facial-features-only model performs nearly as well as Vantage on its own. As age increases, the gap narrows: Vantage scores improve steadily, whereas the facial-features-only model declines modestly, and by the late thirties the combined model remains only marginally above Vantage.

Figure 6 presents an analogous comparison by borrower income and reveals a similar pattern. For the lowest-income borrowers, the combined model outperforms Vantage by roughly 6 percentage points, and the facial-features-only model also carries meaningful signal. As income rises, this incremental gain declines monotonically, falling to under 1 percentage point for the highest earners.

Taken together, these results reinforce a consistent pattern: facial features provide mean-

ingful incremental predictive power primarily for borrowers with limited traditional information, namely those who are younger, lower-income, or have thin credit files. In contrast, for borrowers with richer credit histories, higher incomes, or greater age—where conventional measures already perform well—the additional value of facial features is smaller.

4.3 Facial Features Predicting Vantage Scores

Our analysis so far has used realized delinquency as the outcome. Delinquency, however, is a noisy, binary realization of underlying risk, sensitive to idiosyncratic shocks. Credit scores, by contrast, are designed as ex-ante risk indices, aggregating rich historical information into a smoother measure of expected default risk. We now test whether facial features also predict Vantage scores. This offers a complementary and potentially cleaner signal-extraction exercise: if facial embeddings capture economically meaningful information about borrowers' latent risk, they should help explain variation in these scores as well. Moreover, this test speaks directly to whether facial features encode "hardened soft information," that is, structured proxies for personality traits and economic preferences that traditional scoring models capture only imperfectly.

We implement these tests by estimating equation (1), with the current (i.e., 2019) Vantage score as the outcome, and p_i^{faces} now a prediction of Vantage scores from a similar ML algorithm as for delinquencies, trained as a continuous regression rather than a binary classifier. Table 7 reports the results. In columns (1) to (3) of Panel A, 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. Adding demographics and age controls together with the facial-feature predicted score in column (5) increases the R^2 to 19.1%, indicating 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 again adds human capital controls and shows that facial features remain an important predictor of credit scores. Log salary alone explains 8.5% of variation in Vantage scores, and school rank explains 1.2%; the facial credit score alone explains 11.8%, outperforming both. Combining all three with full demographic controls yields an R^2 of 19.7%. Borrowers 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.

Overall, these results show that facial images contain meaningful information about borrowers' credit risk that is incremental to standard demographic and human capital controls. First, they reinforce our earlier findings based on realized delinquency by demonstrating similarly strong relationships for a smoother, ex-ante measure of risk. Second, they show that facial features systematically align with variation in established credit scores, suggesting that the facial-feature models capture signals closely related to how creditworthiness is assessed in practice. This alignment also suggests that facial features may be particularly useful in settings where traditional credit information is limited or unavailable.

4.4 Robustness: Demographically Homogeneous Subsample

The preceding analyses control for sex and ethnicity through fixed effects but pool borrowers across demographic groups. A natural concern is that facial embeddings partly capture cross-group demographic variation, rather than within-group personality and preference differences. To address this, we replicate our analysis within a single predominant borrower group that is homogeneous in terms of sex and race, eliminating cross-group variation entirely. The results, reported in Appendix E, closely mirror those in the full sample. Facial-feature-based predictions retain similar predictive power for both delinquency and credit scores and the incremental contribution of facial features remains economically meaningful and is again concentrated among borrowers with thin credit files. These findings confirm that our main results are not driven by cross-group demographic differences.

5. Personality and Economic Preferences as Underlying Mechanisms

The results so far establish that facial embeddings predict credit outcomes beyond demographics, income, and education, but do not identify underlying mechanisms. One plausible explanation is that they capture variation in personality traits and economic preferences that shape borrowing and repayment decisions. A large literature in psychology and economics shows that personality traits, especially conscientiousness, are strongly associated with financial behavior (e.g., Almlund et al. (2011), Heckman and Kautz (2012), Falk et al. (2018), Jagelka (2024)), while structural models of household credit emphasize the roles of economic preferences such as patience and risk tolerance (e.g., Chatterjee et al. (2007), Livshits et al. (2007), Meier and Sprenger (2010)). Simultaneously, as discussed in Section 2.4, a growing body of work in genetics and psychology suggests that traits are partially reflected in facial appearance. In this section we test this mechanism by examining whether the photo-based ML measures of Big 5 personality traits and economic preferences introduced in Section 2.4 predict realized delinquency and credit scores, both unconditionally and conditional on traditional credit measures.

Delinquency. Table 8 reports the relationship between realized delinquency and Photo Big 5 personality traits (Panel A) and predicted economic preferences (Panel B). In Panel A, all five personality traits are jointly significant both statistically and economically: the *Top20-Bottom20* differential in column 1 implies that borrowers in the top quintile of predicted personality scores are 4.2 percentage points more likely to become delinquent than those in the bottom quintile, representing a meaningful share of the unconditional mean delinquency rate of 7 percentage points. Personality traits retain their explanatory power after controlling for demographics, age, and the lagged Vantage score (column 4). The *Top20-Bottom20* differential in column 4 remains at 2.8 percentage points. This indicates that personality-based measures capture variation in credit risk beyond standard observables, including lagged credit scores, which may themselves partly reflect underlying traits.

Panel B shows that predicted preferences load on delinquency in the direction expected by economic intuition. Higher predicted risk-taking is associated with significantly greater

delinquency, while greater predicted patience is associated with lower delinquency. These effects survive the inclusion of all controls in column 4, with a *Top20-Bottom20* differential of 2.7 percentage points.

Table 9 splits the sample by credit file thickness. For both the Big 5 (Panel A) and economic preferences (Panel B), the incremental predictive power of personality traits and economic preferences over the lagged Vantage score is larger for thin-file borrowers. In Panel A, the *Top20-Bottom20* differential conditional on the lagged Vantage score rises from 2.4 percentage points for thick-file borrowers (column 3) to 3.6 percentage points for thin-file borrowers (column 6). Panel B exhibits a similar pattern for economic preferences. These results are consistent with our earlier findings: when credit histories are rich, lagged Vantage scores already capture much of the relevant variation, whereas for thin-file borrowers, noisier scores leave more room for personality traits and economic preferences to provide incremental signal.

Vantage Scores. Tables 10 and 11 test whether photo-based personality traits and economic preferences also predict Vantage scores as ex ante measures of borrower risk. In Table 10, both the Big 5 (Panel A) and economic preferences (Panel B) are strongly associated with Vantage scores, with *Top20-Bottom20* differentials of 0.435 and 0.288 standard deviations, respectively, after controlling for demographics. These magnitudes are economically large. Among the economic preferences, the signs are again intuitive: greater risk-taking is associated with lower Vantage scores, while greater patience predicts higher scores. These results show that the photo-based measures capture variation in creditworthiness that is at least partly reflected in traditional scoring models. Put differently, a portion of the information embedded in credit scores – information that is correlated with soft, personality traits and economic preferences – can be codified from facial features into structured data. Our measures thus represent "hardened" soft information: structured, scalable proxies for personality traits and economic preferences that traditional credit scores capture only imperfectly.

Table 11 disaggregates by credit file thickness. Both the Big 5 (Panel A) and economic preferences (Panel B) predict Vantage scores for thick- and thin-file borrowers alike, though the magnitudes are somewhat larger for thick-file borrowers. This pattern is the mirror image of the delinquency results in Table 9: there, the incremental predictive power of

personality traits and economic preferences over the lagged Vantage score is larger for thin-file borrowers while here, personality traits and economic preferences explain more of the variation in Vantage scores for thick-file borrowers because Vantage scores are more precisely estimated for individuals with richer credit histories, and thus better reflect the accumulated behavioral patterns shaped by personality and economic preferences.

Two qualifications are worth noting. First, measurement error in the photo-based scores likely attenuates the estimated coefficients, implying that the true associations may be stronger than our point estimates. Second, the modest R^2 values indicate that stable dispositions explain only a limited share of variation in credit outcomes. Idiosyncratic shocks, such as job loss or health events, which are unrelated to personality and preferences but may also affect facial appearance, likely account for additional variation.

6. Conclusion

Using data on more than one million U.S. borrowers, this paper shows that standardized facial embeddings extracted from LinkedIn profile images contain incremental, out-of-sample predictive signal about consumer creditworthiness, with this information concentrated precisely where traditional credit scores are least informative. One reason facial features are informative is that they convey cues related to underlying personality and economic preferences. Consistent with this, we show that predictions of these traits based on facial features are themselves predictive of credit outcomes.

These findings suggest that facial information encodes economically meaningful "soft" signals that are not fully captured by traditional credit measures and may help reduce measurement error in settings where "hard" information is sparse. At the same time, the use of facial data in lending raises profound questions about privacy, consent, algorithmic transparency, and disparate impact that lie beyond the scope of this paper. Instead, we view our results as documenting an empirical regularity – that faces contain "soft" information about creditworthiness – rather than advocating for any particular application or policy. Whether and how such information should be used remains an open question for regulators, lenders, and society.

References

Almlund, M., A. L. Duckworth, J. Heckman, and T. Kautz (2011). Personality psychology and economics. In E. A. Hanushek, S. Machin, and L. Woessmann (Eds.), *Handbook of the Economics of Education*, Volume 4 of *Handbook of the Economics of Education*, pp. 1–181. Elsevier.

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.

Berg, T., A. Fuster, and M. Puri (2022). FinTech lending. *Annual Review of Financial Economics 14*, 187–207.

Blattner, L. and S. Nelson (2021). How costly is noise? data and disparities in consumer credit. *Working Paper*.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. ter Weel (2008). The economics and psychology of personality traits. *Journal of Human Resources 43*(4), 972–1059.

Brown, S. and K. Taylor (2014). Household finances and the 'big five' personality traits. *Journal of Economic Psychology 45*, 197–212.

Chatterjee, S., D. Corbae, M. Nakajima, and J.-V. Ríos-Rull (2007). A quantitative theory of unsecured consumer credit with risk of default. *Econometrica 75*(6), 1525–1589.

Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data 5*(2), 153–163.

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.

Einav, L., A. Finkelstein, R. Kluender, and P. Schrimpf (2016). Beyond statistics: The economic content of risk scores. *American Economic Journal: Applied Economics 8*(2), 195–224.

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2018). Global evidence on economic preferences. *Quarterly Journal of Economics 133*(4), 1645–1692.

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.

Goodman, L. (2024). Classic FICO versus VantageScore 4.0. Urban Institute Report.

Greenwald, D. L., S. T. Howell, C. Li, and E. Yimfor (2024). Regulatory arbitrage or random errors? Implications of race prediction algorithms in fair lending analysis. *Journal of Financial Economics 157*, 103857.

Guenzel, M., S. Kogan, M. Niessner, and K. Shue (2025). AI personality extraction from faces: Labor market implications. Technical Report SSRN 5089827, SSRN.

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.

Heckman, J. J. and T. Kautz (2012). Hard evidence on soft skills. *Labour Economics 19*(4), 451–464.

Heckman, J. J., J. Stixrud, and S. Urzua (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics 24*(3), 411–482.

Indarte, S., R. Kluender, U. Malmendier, and M. Stepner (2026). Consumption wedges: Measuring and diagnosing distortions. NBER Working Paper No. 34891.

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.

Jagelka, T. (2024). Are economists' preferences psychologists' personality traits? a structural approach. *Journal of Political Economy 132*(3), 910–970.

Kachur, A., E. Osin, D. Davydov, K. Shutilov, and A. Novokshonov (2020). Assessing the big five personality traits using real-life static facial images. *Nature Scientific Reports 10*(1), 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., S. Mullainathan, and M. Raghavan (2016). Inherent trade-offs in the fair determination of risk scores. *arXiv preprint arXiv:1609.05807*.

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.

Liberti, J. M. and M. A. Petersen (2019). Information: Hard and soft. *Review of Corporate Finance Studies 8*(1), 1–41.

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.

Livshits, I., J. MacGee, and M. Tertilt (2007). Consumer bankruptcy: A fresh look. *American Economic Review 97*(1), 400–418.

Ludwig, J. and S. Mullainathan (2024). Machine learning as a tool for hypothesis generation. *The Quarterly Journal of Economics 139*(2), 751–827.

Meier, S. and C. Sprenger (2010). Present-biased preferences and credit card borrowing. *American Economic Journal: Applied Economics 2*(1), 192–210.

Petersen, M. A. and R. G. Rajan (1994). The benefits of lending relationships: Evidence from small business data. *The Journal of Finance 49*(1), 3–37.

Petersen, M. A. and R. G. Rajan (2002). Does distance still matter? the information revolution in small business lending. *The Journal of Finance 57*(6), 2533–2570.

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.

Ravina, E. (2020). Love & loans: The effect of beauty and personal characteristics in credit markets. *Journal of Financial Economics 137*(2), 320–352.

Roberts, B. W., N. R. Kuncel, R. Shiner, A. Caspi, and L. R. Goldberg (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science 2*(4), 313–345.

Ru, H. and A. Schoar (2016). Do credit card companies screen for behavioral biases? NBER Working Paper No. 22360.

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.

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.

Todorov, A., A. N. Mandisodza, A. Goren, and C. C. Hall (2005). Inferences of competence from faces predict election outcomes. *Science 308*(5728), 1623–1626.

Vallée, B. and Y. Zeng (2019). Marketplace lending: A new banking paradigm? *The Review of Financial Studies 32*(5), 1939–1982.

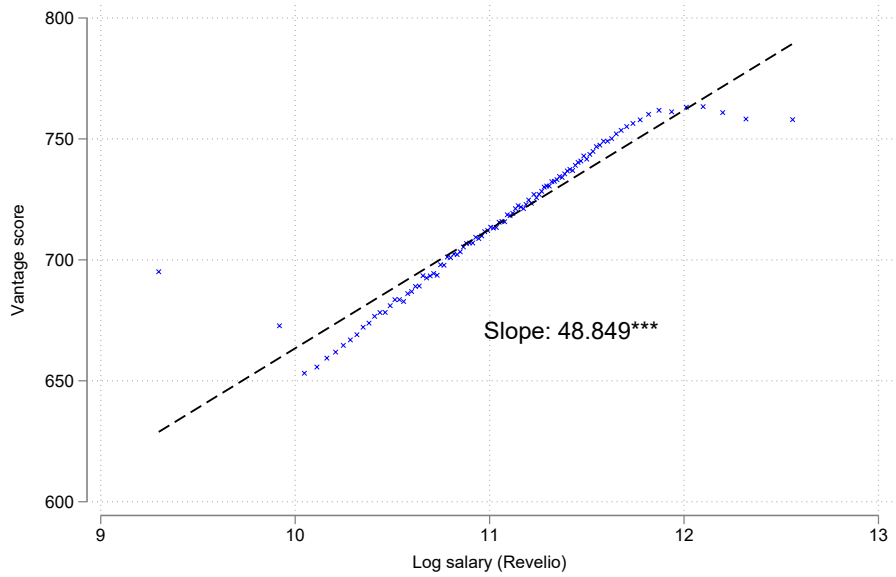
Vukasović, T. and D. Bratko (2015). Heritability of personality: A meta-analysis of behavior genetic studies. *Psychological Bulletin 141*(4), 769–785.

Willis, J. and A. Todorov (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science 17*(7), 592–598.

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

This figure presents binned scatter plots showing the correlations of Vantage scores with (a) log salary and (b) U.S. undergraduate school rank. Salary and school rank are obtained from Revelio Labs data. The dashed lines represent fitted values from a linear regression.

a)



b)

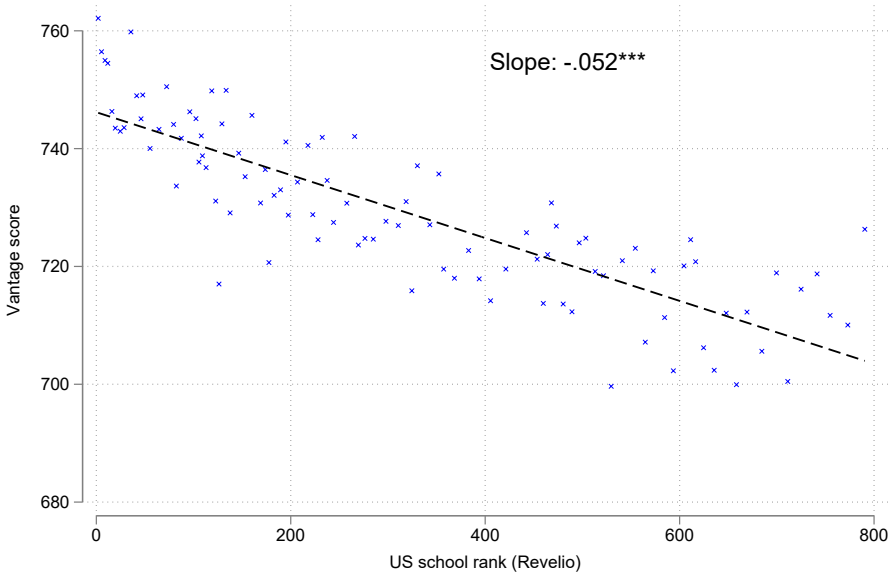


Figure 2: ROC Comparison

This plot shows Receiver Operating Characteristic (ROC) curve to measure the performance of the lagged Vantage score alone (blue) versus the predictions from a ML model combining facial features with lagged Vantage score (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-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 $t-24$.

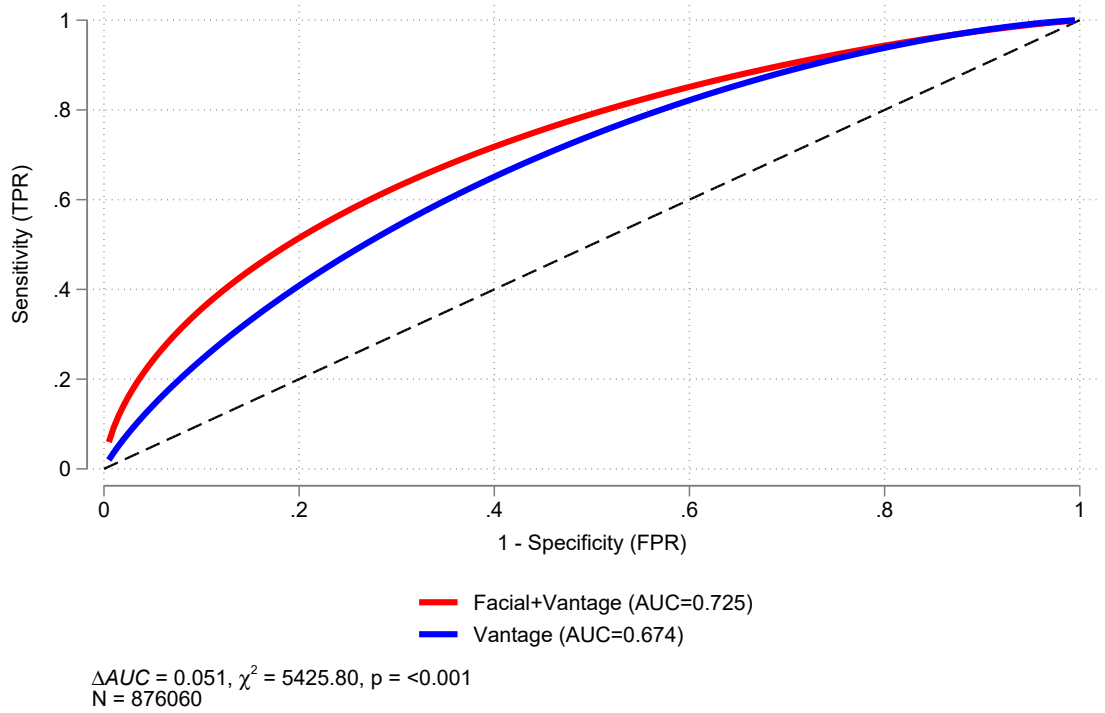


Figure 3: Number of Tradelines

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: a ML model using facial features only (green circles), lagged Vantage score only (blue squares), and a ML model combining facial features with lagged Vantage score (red diamonds). All three logistic regressions are estimated with demographic and age controls. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30+ days-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 $t-24$.

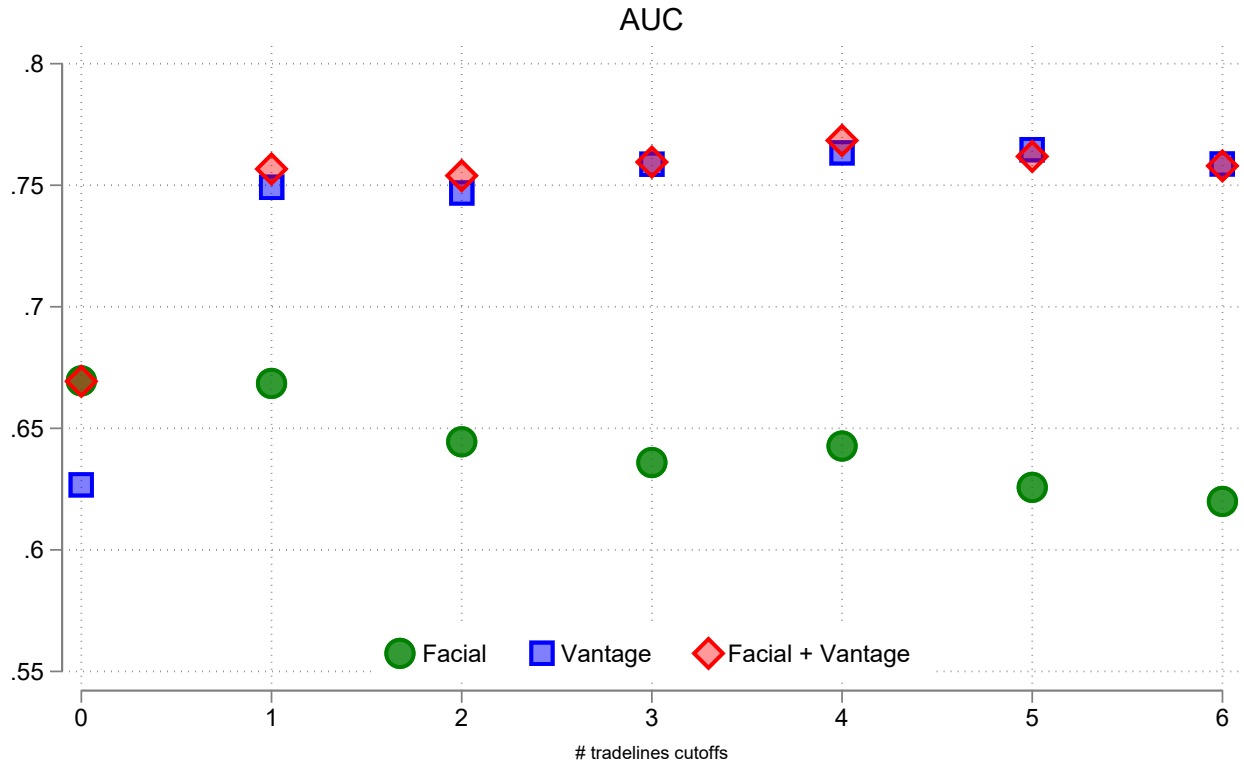


Figure 4: ROC Comparison for Thick versus Thin Files

This plot shows Receiver Operating Characteristic (ROC) curves to measure the performance of the lagged Vantage score alone (blue) versus the predictions from a ML model combining facial features with lagged Vantage score (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-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 $t-24$.

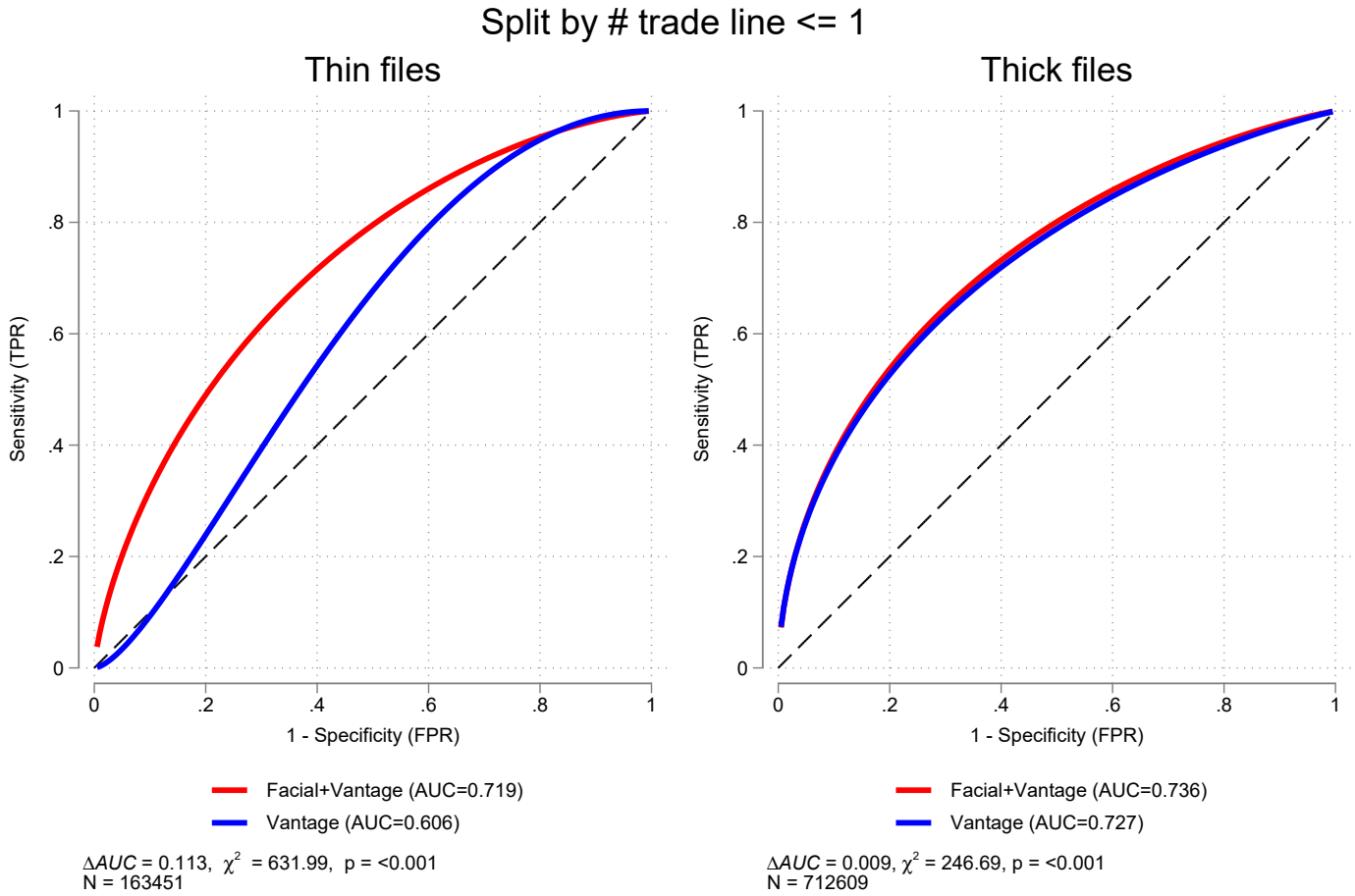


Figure 5: Borrower Age

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: a ML model using facial features only (green circles), lagged Vantage score only (blue squares), and a ML model combining facial features with lagged Vantage score (red diamonds). All three logistic regressions are estimated with demographic and age controls. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30+ days-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 $t-24$.

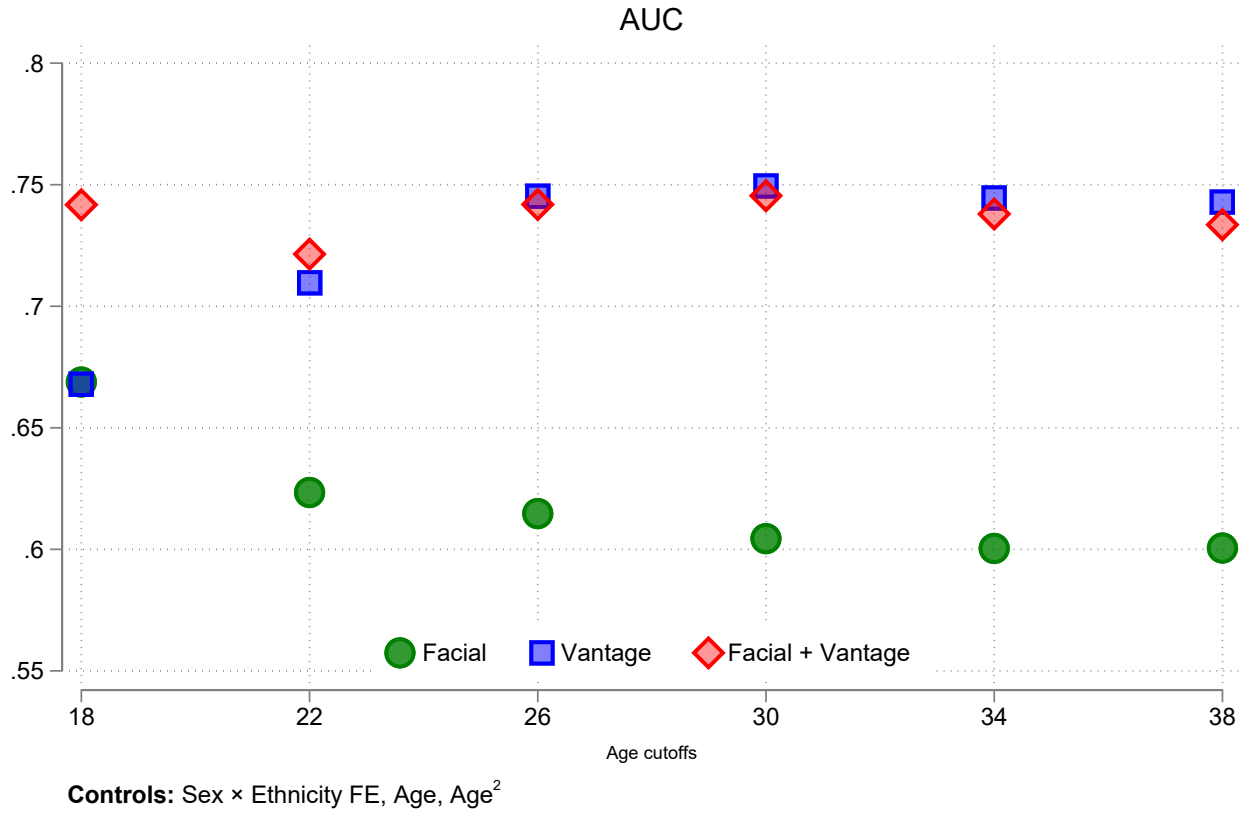


Figure 6: Borrower Income

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: a ML model using facial features only (green circles), lagged Vantage score only (blue squares), and a ML model combining facial features with lagged Vantage score (red diamonds). All three logistic regressions are estimated with demographic and age controls. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30+ days-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 $t-24$.

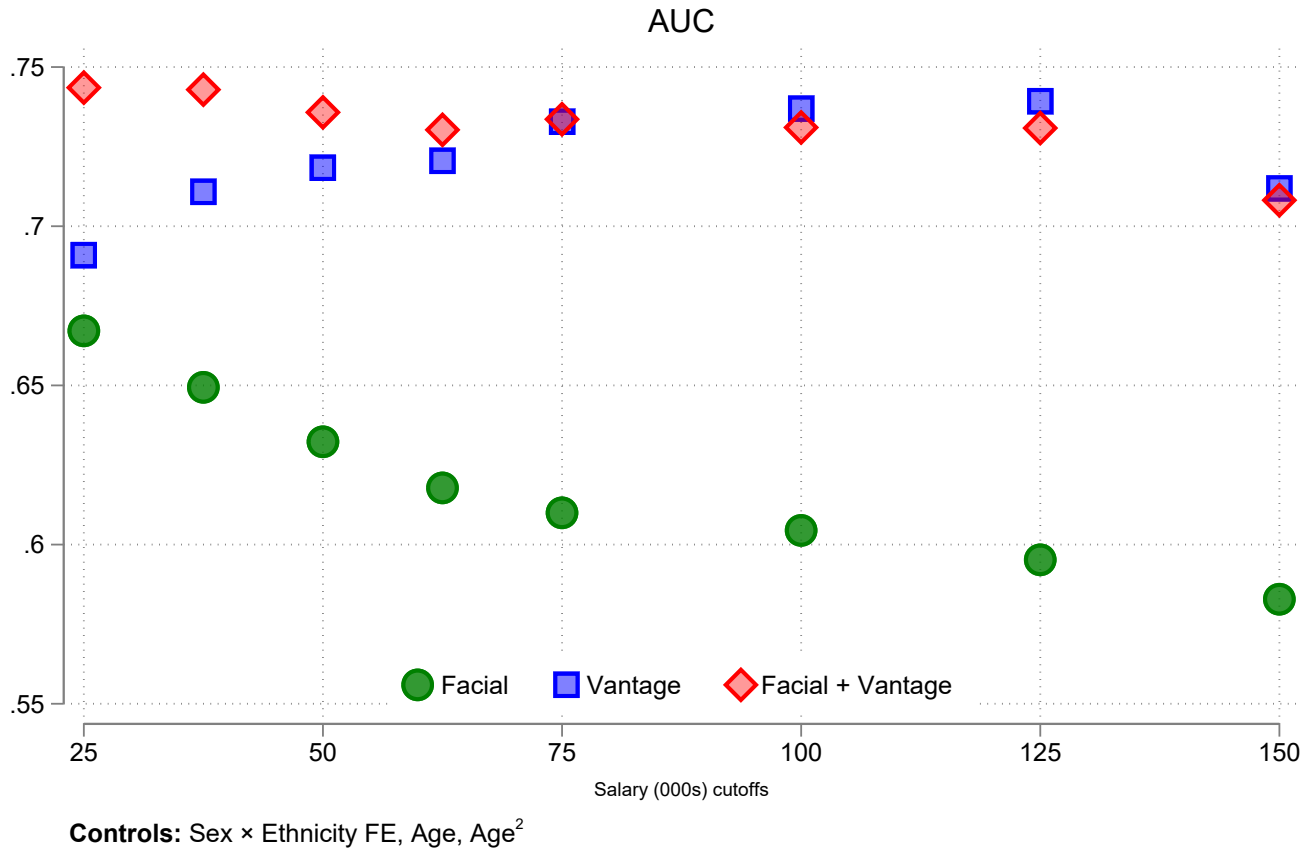


Table 1: Data Description

This table provides summary statistics for key variables used in our analysis. t denotes 2019; lagged variables are measured at $t-2$ (2017). $\mathbb{1}$(Sex = Male) and $\mathbb{1}$(Ethnicity = 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 is the U.S. News ranking of the borrower's undergraduate institution, where lower values indicate higher-ranked schools. Age, Salary and US school rank are computed as of $t-2$ or the earliest available data prior to $t-2$. 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. For each, we report the realized binary indicator and the corresponding continuous prediction from our facial ML model. The first definition identifies borrowers with a 30+ days-past-due episode in the 12 months preceding t, conditional on being current at $t-2$. The second definition identifies borrowers experiencing an increase in 30+ days-past-due tradelines during the same 12-month window, without conditioning on prior current status. Photo Big 5 and economic preferences for risk-taking and patience are predicted from facial photographs using a deep learning model trained on self-reported scores.

	Obs	Mean	Std. Dev.	P10	Median	P90
$\mathbb{1}$(Sex = Male)	1,097,588	0.45	0.50	0.00	0.00	1.00
$\mathbb{1}$(Ethnicity = White)	1,097,588	0.79	0.41	0.00	1.00	1.00
Age	1,097,588	31.28	10.93	20.00	28.00	48.00
Salary ($)	1,047,228	74836.69	44620.67	31724.45	66052.80	123064.75
Thin file ($\leq$ 1 accounts, t-2)	975,663	0.18	0.38	0.00	0.00	1.00
US school rank	666,660	321.80	226.82	46.00	275.00	655.00
Vantage score	1,097,588	713.06	93.28	561.00	729.00	817.00
Facial credit score	1,097,588	704.93	32.51	662.35	707.47	744.23
$\mathbb{1}$(30d+ delinq past 12m & current in t-2)	876,304	0.07	0.25	0.00	0.00	0.00
Facial delinquency (30d+ delinq past 12m & current in t-2)	877,991	0.07	0.02	0.04	0.07	0.10
$\mathbb{1}$(Increase 30d+ delinq past 12m)	964,214	0.07	0.26	0.00	0.00	0.00
Facial delinquency (Increase 30d+ delinq past 12m)	965,177	0.07	0.02	0.05	0.07	0.09
Extraversion	1,097,588	0.54	0.04	0.50	0.54	0.60
Conscientiousness	1,097,588	0.68	0.05	0.62	0.67	0.74
Neuroticism	1,097,588	0.48	0.07	0.39	0.48	0.56
Agreeableness	1,097,588	0.71	0.03	0.69	0.71	0.74
Openness	1,097,588	0.72	0.03	0.68	0.72	0.75
Risk-taking	1,097,588	0.59	0.05	0.53	0.59	0.67
Patience	1,097,588	0.71	0.04	0.67	0.71	0.75

Table 2: Delinquencies Prediction

This table regresses realized delinquencies on the facial delinquencies score, log salary and U.S. undergraduate school rank. Panel A presents specifications including only the facial delinquencies score and demographic fixed effects or age controls. Panel B adds controls for human capital proxies: log salary and U.S. undergraduate school rank. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30+ days-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 × 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***	0.022***
				(0.000)	(0.000)
Sex × 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	876,304	876,304	876,304	876,304	876,304
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***			0.025***	0.021***
		(0.000)			(0.000)	(0.000)
Log salary (Z)			-0.004***		-0.002***	-0.005***
			(0.000)		(0.000)	(0.000)
US school rank (Z)				0.011***	0.008***	0.007***
				(0.000)	(0.000)	(0.000)
No US school rank				0.010***	0.008***	0.006***
				(0.001)	(0.001)	(0.001)
Sex × 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	851,376	851,376	851,376	851,376	851,376	851,376
Top20-Bottom20		0.071			0.068	0.058
AUC	0.591	0.607	0.520	0.544	0.611	0.629

Table 3: Delinquencies Prediction with Lagged Vantage Score

This table regresses realized delinquencies on the facial delinquencies score, log salary, U.S. undergraduate school rank, and lagged Vantage score. Panel A presents specifications including the facial delinquencies score, lagged Vantage 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-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 × 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***		0.022***	0.014***
		(0.000)		(0.000)	(0.000)
Vantage score (Z, t-2)			-0.041***	-0.038***	-0.056***
			(0.000)	(0.000)	(0.000)
Sex × 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	876,060	876,060	876,060	876,060	876,060
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***				0.022***	0.014***
		(0.000)				(0.000)	(0.000)
Log salary (Z)			-0.004***			0.011***	0.001**
			(0.000)			(0.000)	(0.000)
US school rank (Z)				0.011***		0.005***	0.003***
				(0.000)		(0.000)	(0.000)
No US school rank				0.010***		0.006***	0.002***
				(0.001)		(0.001)	(0.001)
Vantage score (Z, t-2)					-0.042***	-0.043***	-0.057***
					(0.000)	(0.000)	(0.000)
Sex × 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	0.033	0.045
Observations	851,143	851,143	851,143	851,143	851,143	851,143	851,143
Top20-Bottom20		0.071				0.060	0.037
AUC	0.591	0.607	0.520	0.544	0.677	0.699	0.724

Table 4: Delinquencies Prediction (Facial + Vantage)

This table regresses realized delinquency on delinquency score predicted from a combined ML model of facial features and lagged Vantage score. Panel A presents specifications benchmarking our combined ML prediction against the lagged Vantage score, controlling for demographic fixed effects and age. Panel B adds controls for log salary and U.S. undergraduate school rank. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30+ days-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 × ethnicity fixed effects and age controls are included where indicated. *Top20–Bottom20* is the difference between the average predicted delinquency score of the top and bottom quintile of individuals based on the facial features + Vantage-based delinquency 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 + Vantage delinquency (Z)		0.057***	0.058***		
		(0.000)	(0.000)		
Vantage score (Z, t-2)				-0.041***	-0.059***
				(0.000)	(0.000)
Sex × Ethnicity FE	Yes	No	Yes	No	Yes
Age + Age^2	Yes	No	Yes	No	Yes
R2	0.007	0.050	0.056	0.023	0.042
Observations	876,060	876,060	876,060	876,060	876,060
Top20-Bottom20		0.148	0.151		
AUC	0.591	0.723	0.735	0.677	0.716

Panel B

	Delinquency				
	(1)	(2)	(3)	(4)	(5)
Facial + Vantage delinquency (Z)	0.057***			0.058***	0.058***
	(0.000)			(0.000)	(0.000)
Log salary (Z)		-0.004***		0.007***	-0.001**
		(0.000)		(0.000)	(0.000)
US school rank (Z)			0.011***	0.004***	0.003***
			(0.000)	(0.000)	(0.000)
No US school rank			0.010***	0.005***	0.002***
			(0.001)	(0.001)	(0.001)
Sex × Ethnicity FE	No	No	No	No	Yes
Age + Age^2	No	No	No	No	Yes
R2	0.051	0.000	0.002	0.051	0.056
Observations	851,143	851,143	851,143	851,143	851,143
Top20-Bottom20	0.149			0.151	0.151
AUC	0.723	0.520	0.544	0.724	0.735

Table 5: Thick versus Thin Files

This table regresses realized delinquencies on the facial delinquencies score from our ML predictions based on facial features alone and lagged Vantage score. Analysis is conducted separately by credit file thickness. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30+ days-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 fixed effects and age controls are included 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)	0.025***		0.013***	0.015***		0.013***
	(0.000)		(0.000)	(0.000)		(0.000)
Vantage score (Z, t-2)		-0.080***	-0.077***		-0.017***	-0.015***
		(0.000)	(0.000)		(0.001)	(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	712,609	712,609	712,609	163,451	163,451	163,451
Top20-Bottom20	0.067		0.035	0.043		0.038
AUC	0.610	0.741	0.744	0.668	0.673	0.696
E(Y)		0.075			0.038	

Table 6: Delinquencies Prediction by Credit File Thickness (Facial + Vantage)

This table regresses realized delinquency on the delinquency score predicted from a combined ML model, benchmarking it against the lagged Vantage score. Analysis is conducted separately by credit file thickness. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30+ days-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 fixed effects and age controls are included 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 quintile of individuals based on the facial features + Vantage-based delinquencies score. 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)
Facial + Vantage delinquency (Z)	0.067***		0.026***	
	(0.000)		(0.000)	
Vantage score (Z, t-2)		-0.080***		-0.017***
		(0.000)		(0.001)
Sex × Ethnicity FE	Yes	Yes	Yes	Yes
Age + Age^2	Yes	Yes	Yes	Yes
R2	0.064	0.058	0.028	0.015
Observations	712,609	712,609	163,451	163,451
Top20-Bottom20	0.171		0.074	
AUC	0.737	0.741	0.728	0.673
E(Y)	0.075		0.038	

Table 7: 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 human capital proxies: log salary and U.S. undergraduate school rank. Variables denoted with (Z) are standardized to have mean 0 and standard deviation 1. Sex × 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***	0.199***
				(0.001)	(0.001)
Sex × 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	1,097,588	1,097,588	1,097,588	1,097,588	1,097,588
Top20-Bottom20				0.989	0.553

Panel B

	Vantage score (Z)					
	(1)	(2)	(3)	(4)	(5)	(6)
Facial credit score (Z)		0.345***			0.278***	0.181***
		(0.001)			(0.001)	(0.001)
Log salary (Z)			0.289***		0.195***	0.128***
			(0.001)		(0.001)	(0.001)
US school rank (Z)				-0.125***	-0.070***	-0.078***
				(0.001)	(0.001)	(0.001)
No US school rank				-0.091***	-0.096***	-0.105***
				(0.002)	(0.002)	(0.002)
Sex × Ethnicity FE	Yes	No	No	No	No	Yes
Age + Age^2	Yes	No	No	No	No	Yes
R2	0.149	0.119	0.085	0.012	0.160	0.197
Observations	1,047,228	1,047,228	1,047,228	1,047,228	1,047,228	1,047,228
Top20-Bottom20		0.946			0.763	0.495

Table 8: Delinquency Prediction from Personality Traits and Economic Preferences

This table regresses delinquency against their personality traits (Photo Big 5) and economic preferences (risk-taking and patience), and lagged Vantage score. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30+ days-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 fixed effects and age controls are included 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 delinquency of the top and bottom quintile of individuals based on their combined scores for personality traits (Panel A) and economic preferences (Panel B). 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)
Openness (Z)	0.015***	0.016***		0.009***
	(0.000)	(0.000)		(0.000)
Conscientiousness (Z)	-0.004***	-0.006***		-0.004***
	(0.001)	(0.001)		(0.001)
Extraversion (Z)	0.007***	0.005***		0.005***
	(0.000)	(0.000)		(0.000)
Agreeableness (Z)	0.002**	-0.004***		-0.003***
	(0.001)	(0.001)		(0.001)
Neuroticism (Z)	-0.009***	-0.013***		-0.008***
	(0.000)	(0.001)		(0.001)
Vantage score (Z, t-2)			-0.055***	-0.054***
			(0.000)	(0.000)
Sex $\times$ Ethnicity FE	No	Yes	Yes	Yes
Age + Age^2	No	Yes	Yes	Yes
R2	0.005	0.010	0.042	0.043
Observations	899,959	899,959	899,959	899,959
Top20-Bottom20	0.042	0.046		0.028
AUC	0.570	0.607	0.716	0.718

Panel B

	Delinquency			
	(1)	(2)	(3)	(4)
Risk-taking (Z)	0.011***	0.018***		0.012***
	(0.000)	(0.000)		(0.000)
Patience (Z)	-0.003***	-0.005***		-0.004***
	(0.000)	(0.000)		(0.000)
Vantage score (Z, t-2)			-0.055***	-0.054***
			(0.000)	(0.000)
Sex $\times$ Ethnicity FE	No	Yes	Yes	Yes
Age + Age^2	No	Yes	Yes	Yes
R2	0.001	0.009	0.042	0.043
Observations	899,959	899,959	899,959	899,959
Top20-Bottom20	0.024	0.039		0.027
AUC	0.536	0.602	0.716	0.718

Table 9: Delinquency Prediction from Personality Traits and Economic Preferences by Credit File Thickness

This table regresses delinquency against their personality traits (Photo Big 5) and economic preferences (risk-taking and patience), and lagged Vantage score. Analysis is conducted separately by credit file thickness. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30+ days-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 × ethnicity fixed effects and age controls are included 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 delinquency of the top and bottom quintile of individuals based on their combined scores for personality traits (Panel A) and economic preferences (Panel B). 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)	(5)	(6)
Openness (Z)	0.017***		0.007***	0.013***		0.011***
	(0.000)		(0.000)	(0.001)		(0.001)
Conscientiousness (Z)	-0.006***		-0.003***	-0.002*		-0.002
	(0.001)		(0.001)	(0.001)		(0.001)
Extraversion (Z)	0.005***		0.005***	0.002***		0.002***
	(0.001)		(0.001)	(0.001)		(0.001)
Agreeableness (Z)	-0.004***		-0.003***	-0.003**		-0.003**
	(0.001)		(0.001)	(0.001)		(0.001)
Neuroticism (Z)	-0.014***		-0.007***	-0.012***		-0.011***
	(0.001)		(0.001)	(0.001)		(0.001)
Vantage score (Z, t-2)		-0.080***	-0.079***		-0.017***	-0.016***
		(0.000)	(0.000)		(0.001)	(0.001)
Sex × Ethnicity FE	Yes	Yes	Yes	Yes	Yes	Yes
Age + Age^2	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.009	0.058	0.059	0.014	0.015	0.019
Observations	732,761	732,761	732,761	167,198	167,198	167,198
Top20-Bottom20	0.048		0.024	0.041		0.036
AUC	0.586	0.741	0.742	0.655	0.673	0.688
E(Y)		0.076			0.038	

Panel B

	Delinquency					
	Thick file			Thin file		
	(1)	(2)	(3)	(4)	(5)	(6)
Risk-taking (Z)	0.019***		0.011***	0.013***		0.011***
	(0.001)		(0.001)	(0.001)		(0.001)
Patience (Z)	-0.006***		-0.003***	-0.001		-0.000
	(0.000)		(0.000)	(0.001)		(0.001)
Vantage score (Z, t-2)		-0.080***	-0.080***		-0.017***	-0.017***
		(0.000)	(0.000)		(0.001)	(0.001)
Sex × Ethnicity FE	Yes	Yes	Yes	Yes	Yes	Yes
Age + Age^2	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.007	0.058	0.058	0.012	0.015	0.018
Observations	732,761	732,761	732,761	167,198	167,198	167,198
Top20-Bottom20	0.041		0.024	0.034		0.030
AUC	0.581	0.741	0.742	0.648	0.673	0.685
E(Y)		0.076			0.038	

Table 10: Vantage Score Prediction from Personality Traits and Economic Preferences

This table regresses Vantage scores against their personality traits (Photo Big 5) and economic preferences (risk-taking and patience). Sex × ethnicity fixed effects and age controls are included 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 Vantage score of the top and bottom quintile of individuals based on their combined scores for personality traits (Panel A) and economic preferences (Panel B). 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)
Openness (Z)	-0.206***	-0.157***
	(0.001)	(0.001)
Conscientiousness (Z)	-0.010***	0.034***
	(0.003)	(0.002)
Extraversion (Z)	-0.165***	-0.003**
	(0.002)	(0.001)
Agreeableness (Z)	0.232***	0.026***
	(0.002)	(0.002)
Neuroticism (Z)	-0.101***	0.125***
	(0.002)	(0.002)
Sex × Ethnicity FE	No	Yes
Age + Age^2	No	Yes
R2	0.037	0.184
Observations	1,097,588	1,097,588
Top20-Bottom20	0.511	0.435

Panel B

	Vantage score (Z)	
	(1)	(2)
Risk-taking (Z)	-0.236***	-0.129***
	(0.001)	(0.002)
Patience (Z)	0.186***	0.033***
	(0.001)	(0.001)
Sex × Ethnicity FE	No	Yes
Age + Age^2	No	Yes
R2	0.023	0.171
Observations	1,097,588	1,097,588
Top20-Bottom20	0.412	0.288

Table 11: Vantage Score Prediction from Personality Traits and Economic Preferences by Credit File Thickness

This table regresses vantage score against their personality traits (Photo Big 5) and economic preferences (risk-taking and patience). Analysis is conducted separately by credit file thickness. 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 fixed effects and age controls are included 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 Vantage score of the top and bottom quintile of individuals based on their combined scores for personality traits (Panel A) and economic preferences (Panel B). 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)	
	Thick file	Thin file
	(1)	(2)
Openness (Z)	-0.169***	-0.134***
	(0.001)	(0.003)
Conscientiousness (Z)	0.040***	0.035***
	(0.003)	(0.006)
Extraversion (Z)	-0.009***	0.000
	(0.002)	(0.004)
Agreeableness (Z)	0.037***	0.004
	(0.003)	(0.006)
Neuroticism (Z)	0.129***	0.130***
	(0.002)	(0.005)
Sex $\times$ Ethnicity FE	Yes	Yes
Age + Age^2	Yes	Yes
R2	0.108	0.057
Observations	827,673	175,215
Top20-Bottom20	0.459	0.405
E(Y)	732.997	665.262

Panel B

	Vantage score (Z)	
	Thick file	Thin file
	(1)	(2)
Risk-taking (Z)	-0.148***	-0.112***
	(0.002)	(0.004)
Patience (Z)	0.047***	0.009**
	(0.002)	(0.004)
Sex $\times$ Ethnicity FE	Yes	Yes
Age + Age^2	Yes	Yes
R2	0.093	0.044
Observations	827,673	175,215
Top20-Bottom20	0.314	0.291
E(Y)	732.997	665.262

Online Appendix
Facing Default?

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

A. Data Cleaning and Merging

This appendix describes the construction of the Revelio-L2 matched sample in detail, covering the cleaning of Revelio data (Section A.1), the pre-merge preparation of L2 data (Section A.2), and the fuzzy matching procedure (Section A.3).

A.1 Cleaning Revelio Data

The unique identifier in Revelio is `user_id`. Each individual is assigned a single `user_id` regardless of relocation across states. Revelio contains three data directories: Profiles (a cross-sectional snapshot of all LinkedIn users), Positions (an individual-by-time panel of employment histories), and Education (an individual-by-time panel of academic histories). Our Revelio vintage is 2023-10-30. We organize the Revelio data into state-level panels to match the state-level structure of the L2 data files.[1]

We apply the following filters to the Revelio data:

1. We retain only U.S.-based profiles, positions, and education records. Because we match to the panel of registered U.S. voters, this restriction is appropriate: individuals must be U.S. citizens or naturalized to appear in L2.[2]

2. We drop observations with missing location or name, and extract the state from the location string.

3. We drop profiles with fewer than 25 LinkedIn connections. Individuals below this threshold often have missing or incomplete information. This step removes approximately 25% of the Revelio sample.

4. Within both the Positions and Education panels, we drop records with missing start dates, periods of stay shorter than 90 days, or missing institution/employer names. From the start date of the first undergraduate degree, we estimate each individual's birth year as the degree start year minus 18.

[1] If a `user_id` relocated from Texas to Michigan in March 2012, they are assigned to the Texas panel prior to March 2012 and to the Michigan panel afterward.

[2] This restriction precludes undergraduate degree information for immigrants who studied abroad and were later naturalized, which limits our ability to estimate their birth year. The number of such cases in the L2 data is likely small.

5. We standardize names by: trimming whitespace; removing content within parentheses, brackets, and braces; removing content following commas (e.g., suffixes such as "PhD"); removing special characters; stripping "junior," "senior," "jr," and "sr" from the start or end of the string; dropping single letters surrounded by spaces; dropping names starting with or containing "the"; removing all non-Latin alphabet strings; and dropping names with fewer than two distinct words.

A.2 Preparing L2 Data and Pre-Merge Filtering

The unique identifier in L2 is `lalvoterid`. Each voter is assigned a unique `lalvoterid` by state, so the identifier is unique at the person-state level. A `lalvoterid` is never reassigned. Each L2 data file is a state-level cross-section of all registered voters as of a given snapshot date.

For each `lalvoterid`, we extract the full name and date of birth. We assign the state based on the state-level file in which the individual appears, with the date of presence set to the vintage date of that file. We append annual vintages for every state to create a single repeated cross-section per state, with unique observations identified by `lalvoterid`, state, and vintage year.

Before merging, we filter both datasets at the state level. From L2, we retain only observations with nonmissing name and nonmissing date of birth. From Revelio, we retain only observations with nonmissing name, nonmissing start and end dates for every position or academic degree in that state, and nonmissing birth year. Across states, approximately 39–45% of Revelio observations have nonmissing birth years; individuals without an estimable birth year are dropped at this stage, as approximate age is required for matching.

A.3 Fuzzy Matching Procedure

The merge is conducted state by state in three steps: fuzzy name matching, date and age filtering, and validation to remove false positives.

Step 1: Fuzzy name matching. For each state, we retain all unique (`lalvoterid`, L2 name) and (`user_id`, Revelio name) pairs. When multiple names exist for a given identifier, we keep the longest string. We then conduct a fuzzy merge between L2 and Revelio names using the Weighted Jaccard Similarity algorithm, implemented via the `fedmatch` R package.[3] We set the similarity threshold at 0.1, chosen through trial and error on randomly selected subsamples to balance precision and recall.

Step 2: Date and age filtering. For each matched name pair, we retain only pairs where: (i) the date ranges of state presence in L2 and Revelio overlap within ±365 days, and (ii) birth years agree within ±2 years. Within each state, we sort matched pairs by descending similarity score for each `user_id` and retain only the highest-scoring match. Because the

[3]The Weighted Jaccard Similarity tokenizes each name into individual words, assigns frequency-based weights to each token, and computes the ratio of the weighted intersection to the weighted union of the two token sets. The measure ranges from 0 (no similarity) to 1 (identical sets). See the `fedmatch` package documentation for details.

procedure loops over states, a single individual may match to different `lalvoterid` records across states.

Step 3: Validation. We apply a second round of LLM-assisted validation to the matched sample, designed to remove nearly all remaining false positives. This step employs rule-based string matching and proceeds as follows. We classify a match as valid if the names agree exactly, with two permitted deviations: at most one spelling error per every seven letters in the name, and missing or abridged middle names. All other cases are classified as nonmatches. Additionally, we compute string similarity scores using several alternative algorithms (Levenshtein, cosine, bag, and MRA normalized similarity) as benchmarks on merge quality, though these do not enter the matching criteria.

B. Photo Big 5 and Preference Prediction: Model Details

This appendix provides additional details on the model used to predict Big 5 personality traits and economic preferences from facial photographs, as described in Section 2.4.

B.1 Architecture

The prediction model is an attention-gated regression network that takes as input the 512-dimensional, L2-normalized facial embedding extracted by the frozen InceptionResNetV1 backbone. The architecture consists of two components.

The first is a shared block, common to all seven outputs, that maps the 512-dimensional input to a 128-dimensional hidden representation via a linear layer, followed by layer normalization, a GELU activation, and dropout ($p = 0.5$).

The second component consists of seven parallel, trait-specific heads. Each head applies a learned gating mechanism: a sigmoid gate produces a 128-dimensional vector of weights in $(0, 1)$ that is multiplied element-wise with the shared representation. This gated representation is then passed through a second linear layer (128 to 128), layer normalization, GELU activation, dropout ($p = 0.5$), and a final linear layer that produces a single scalar output. At inference, outputs are clipped to $[0, 1]$. The gating mechanism allows each head to load on a different combination of the shared features, enabling the model to learn trait-specific and preference-specific patterns in facial appearance.

Table B1 reports the key hyperparameters, all selected via automated hyperparameter search using Optuna.

Table B1: Personality Model Hyperparameters

Hyperparameter	Value
Batch size	32
Learning rate	2.81×10^{-4}
Weight decay	2.66×10^{-4}
Huber δ	0.25
Gradient clip (max norm)	1.0
Warmup epochs	5
Early-stopping patience	15

B.2 Loss Function

The model is trained using an uncertainty-weighted multi-task loss. For each trait $t \in \{1, \ldots, 7\}$, the per-trait loss is the Huber loss (with $\delta = 0.25$), which is robust to outliers in self-reported and single-item survey responses. The total loss is:

$$\mathcal{L}_{\text{total}} = \sum_{t=1}^{7} \left[\exp(-s_t) \cdot \mathcal{L}_t^{\text{Huber}} + s_t \right] \tag{4}$$

where $s_t = \log \sigma_t^2$ is a learnable per-trait log-variance parameter. This formulation automatically down-weights noisier traits and up-weights more reliably measured ones, avoiding the need to hand-tune loss weights across the seven targets.

B.3 Training Procedure

We train with AdamW (learning rate 2.81×10^{-4}, weight decay 2.66×10^{-4}) and gradient clipping at a global norm of 1.0. The learning rate is linearly warmed up over the first 5 epochs (starting at $0.1\times$ the full rate) and subsequently reduced by a factor of 0.5 when the training loss plateaus for 5 consecutive epochs. Training halts when validation loss does not improve for 15 epochs, and the checkpoint with the lowest validation loss is retained. We use five-fold cross-validation: in each fold, 80% of the training pool is used for training and 20% for validation. The final prediction for each individual is the mean across the five fold-specific model outputs.

B.4 Out-of-Sample Performance

Table B2 reports out-of-sample prediction performance on the held-out test set (N = 1,303). We report the correlation between predicted and self-reported scores and R^2.

Table B2: Out-of-Sample Prediction Performance

Trait	Correlation	R^2
Extraversion	0.168	0.026
Conscientiousness	0.213	0.039
Neuroticism	0.244	0.056
Agreeableness	0.142	0.016
Openness	0.130	0.010
Risk-taking	0.229	0.043
Patience	0.142	0.010
Mean	0.181	0.029

The R^2 values are modest, reflecting the fact that personality and preferences are inherently multi-determined: genetic, environmental, and situational factors all contribute to self-reported scores, and a single photograph captures only a fraction of this variation. Nevertheless, the correlations between predicted and self-reported scores (0.13 to 0.24) are comparable to correlations between behavioral task measures and survey-based personality assessments, which rarely exceed 0.3 (Almlund et al., 2011). Measurement error in the predicted scores will attenuate our downstream regression coefficients, implying that the true associations between personality, preferences, and credit outcomes are likely stronger than what we estimate.

C. Appendix Tables

Table C1: 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 TPR 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-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 $t-2$. Differences are estimated via binormal ROC regression with 100 bootstrap repetitions. Standard errors are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

	Full sample	Thin files	Thick files
	(1)	(2)	(3)
FPR = 0.10	0.113***	0.226***	0.006***
	(0.002)	(0.006)	(0.002)
FPR = 0.30	0.088***	0.222***	0.012***
	(0.001)	(0.007)	(0.001)
FPR = 0.50	0.047***	0.118***	0.012***
	(0.001)	(0.006)	(0.001)
FPR = 0.70	0.015***	0.030***	0.010***
	(0.001)	(0.004)	(0.001)
FPR = 0.90	-0.001**	-0.004***	0.004***
	(0.000)	(0.001)	(0.001)
Observations	876060	163451	712609

D. Robustness: Alternate definition of delinquency

Table D1: Delinquencies Prediction

This table regresses realized 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 and lagged Vantage score, controlling for demographic fixed effects and age. 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-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 × 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***		0.020***	0.013***
		(0.000)		(0.000)	(0.000)
Vantage score (Z, t-2)			-0.037***	-0.034***	-0.045***
			(0.000)	(0.000)	(0.000)
Sex × 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	963,956	963,956	963,956	963,956	963,956
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***				0.020***	0.013***
		(0.000)				(0.000)	(0.000)
Log salary (Z)			-0.004***			0.009***	0.000
			(0.000)			(0.000)	(0.000)
US school rank (Z)				0.011***		0.005***	0.004***
				(0.000)		(0.000)	(0.000)
No US school rank				0.011***		0.006***	0.002***
				(0.001)		(0.001)	(0.001)
Vantage score (Z, t-2)					-0.038***	-0.038***	-0.045***
					(0.000)	(0.000)	(0.000)
Sex × 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	937,924	937,924	937,924	937,924	937,924	937,924	937,924
Top20-Bottom20		0.067				0.054	0.035
AUC	0.590	0.601	0.518	0.543	0.669	0.689	0.711

Table D2: Delinquencies Prediction (Facial + Vantage)

This table regresses realized delinquencies on predictions from a combined ML model of facial features and lagged Vantage score, controlling for log salary, U.S. undergraduate school rank, and lagged Vantage score. Panel A presents specifications including the predicted delinquencies score and lagged Vantage score, controlling for demographic fixed effects and age. Panel B adds controls for log salary and U.S. undergraduate school rank. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences an increase in the number of tradelines that are 30+ days-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 × 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 features + Vantage-based 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 + Vantage delinquency (Z)		0.054***	0.054***		
		(0.000)	(0.000)		
Vantage score (Z, t-2)				-0.037***	-0.048***
				(0.000)	(0.000)
Sex × Ethnicity FE	Yes	No	Yes	No	Yes
Age + Age^2	Yes	No	Yes	No	Yes
R2	0.007	0.044	0.049	0.020	0.034
Observations	963,956	963,956	963,956	963,956	963,956
Top20-Bottom20		0.143	0.143		
AUC	0.590	0.715	0.727	0.669	0.705

Panel B

	Delinquency				
	(1)	(2)	(3)	(4)	(5)
Facial + Vantage delinquency (Z)	0.054***			0.055***	0.054***
	(0.000)			(0.000)	(0.000)
Log salary (Z)		-0.004***		0.006***	-0.001**
		(0.000)		(0.000)	(0.000)
US school rank (Z)			0.011***	0.004***	0.003***
			(0.000)	(0.000)	(0.000)
No US school rank			0.011***	0.005***	0.002***
			(0.001)	(0.001)	(0.001)
Sex × Ethnicity FE	No	No	No	No	Yes
Age + Age^2	No	No	No	No	Yes
R2	0.044	0.000	0.002	0.045	0.049
Observations	937,924	937,924	937,924	937,924	937,924
Top20-Bottom20	0.144			0.145	0.143
AUC	0.715	0.518	0.543	0.716	0.726

Table D3: Thick versus Thin Files

This table regresses realized delinquencies on the facial delinquencies score from our ML predictions based on facial features alone and lagged Vantage score, separately for thin and thick-file borrowers. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences an increase in the number of tradelines that are 30+ days-past-due in the 12 months immediately preceding t (the window $[t-12, t]$). In Panel A we regress the deliquency 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 × ethnicity fixed effects and, where indicated, age controls. 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***		0.015***
		(0.000)		(0.000)
Sex × Ethnicity FE	Yes	Yes	Yes	Yes
Age + Age^2	Yes	Yes	Yes	Yes
R2	0.005	0.011	0.009	0.015
Observations	798,387	798,387	165,569	165,569
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***		0.012***	0.015***		0.013***
	(0.000)		(0.000)	(0.000)		(0.000)
Vantage score (Z, t-2)		-0.059***	-0.056***		-0.017***	-0.015***
		(0.000)	(0.000)		(0.001)	(0.001)
Sex × 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	798,387	798,387	798,387	165,569	165,569	165,569
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 D4: Delinquencies Prediction by Credit File Thickness (Facial + Vantage)

This table regresses realized delinquencies on delinquencies score predicted from a combined ML model and lagged Vantage score alone, separately for thin and thick-filed borrowers. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences an increase in the number of tradelines that are 30+ days-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 $t-2$. Variables denoted with (Z) are standardized to have mean 0 and standard deviation 1. Sex × 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 features + Vantage-based delinquencies score. 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)
Facial + Vantage delinquency (Z)	0.060***		0.026***	
	(0.000)		(0.000)	
Vantage score (Z, t-2)		-0.059***		-0.017***
		(0.000)		(0.001)
Sex × Ethnicity FE	Yes	Yes	Yes	Yes
Age + Age^2	Yes	Yes	Yes	Yes
R2	0.052	0.041	0.026	0.015
Observations	798,387	798,387	165,569	165,569
Top20-Bottom20	0.159		0.069	
AUC	0.725	0.721	0.722	0.671
E(Y)	0.079		0.038	

Table D5: 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 TPR 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-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 $t-2$. Differences are estimated via binormal ROC regression with 100 bootstrap repetitions. Standard errors are reported in parentheses. $^*p < 0.10$, $p < 0.05$, $^*p < 0.01$.

	Full sample	Thin files	Thick files
	(1)	(2)	(3)
FPR = 0.10	0.112***	0.200***	0.032***
	(0.002)	(0.007)	(0.002)
FPR = 0.30	0.092***	0.213***	0.026***
	(0.001)	(0.007)	(0.001)
FPR = 0.50	0.048***	0.120***	0.015***
	(0.001)	(0.005)	(0.001)
FPR = 0.70	0.014***	0.035***	0.006***
	(0.001)	(0.003)	(0.001)
FPR = 0.90	-0.001***	-0.002*	0.001*
	(0.000)	(0.001)	(0.000)
Observations	963956	165569	798387

Table D6: Delinquency Prediction from Personality Traits and Economic Preferences

This table regresses delinquency against their personality traits (Photo Big 5) and economic preferences (risk-taking and patience), 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-past-due in the 12 months immediately preceding t (the window $[t-12, t]$). Sex × 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 scores for personality traits (Panel A) and economic preferences (Panel B). 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)
Openness (Z)	0.014***	0.015***		0.008***
	(0.000)	(0.000)		(0.000)
Conscientiousness (Z)	-0.004***	-0.005***		-0.004***
	(0.001)	(0.001)		(0.001)
Extraversion (Z)	0.006***	0.005***		0.004***
	(0.000)	(0.000)		(0.000)
Agreeableness (Z)	0.003***	-0.003***		-0.002***
	(0.001)	(0.001)		(0.001)
Neuroticism (Z)	-0.009***	-0.013***		-0.008***
	(0.000)	(0.001)		(0.001)
Vantage score (Z, t-2)			-0.047***	-0.046***
			(0.000)	(0.000)
Sex × Ethnicity FE	No	Yes	Yes	Yes
Age + Age^2	No	Yes	Yes	Yes
R2	0.004	0.010	0.034	0.035
Observations	990,768	990,768	990,768	990,768
Top20-Bottom20	0.042	0.044		0.026
AUC	0.569	0.605	0.705	0.706

Panel B

	Delinquency			
	(1)	(2)	(3)	(4)
Risk-taking (Z)	0.010***	0.017***		0.011***
	(0.000)	(0.000)		(0.000)
Patience (Z)	-0.002***	-0.005***		-0.003***
	(0.000)	(0.000)		(0.000)
Vantage score (Z, t-2)			-0.047***	-0.047***
			(0.000)	(0.000)
Sex × Ethnicity FE	No	Yes	Yes	Yes
Age + Age^2	No	Yes	Yes	Yes
R2	0.001	0.009	0.034	0.035
Observations	990,768	990,768	990,768	990,768
Top20-Bottom20	0.022	0.038		0.025
AUC	0.532	0.600	0.705	0.706

Table D7: Delinquency Prediction from Personality Traits and Economic Preferences by Credit File Thickness

This table regresses delinquency against their personality traits (Photo Big 5) and economic preferences (risk-taking and patience), and lagged Vantage score. Analysis is conducted separately 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-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 scores for personality traits (Panel A) and economic preferences (Panel B). 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)	(5)	(6)
Openness (Z)	0.016***		0.007***	0.012***		0.011***
	(0.000)		(0.000)	(0.001)		(0.001)
Conscientiousness (Z)	-0.005***		-0.003***	-0.002*		-0.002
	(0.001)		(0.001)	(0.001)		(0.001)
Extraversion (Z)	0.005***		0.004***	0.002**		0.002**
	(0.001)		(0.000)	(0.001)		(0.001)
Agreeableness (Z)	-0.003***		-0.001*	-0.003**		-0.003**
	(0.001)		(0.001)	(0.001)		(0.001)
Neuroticism (Z)	-0.013***		-0.007***	-0.012***		-0.011***
	(0.001)		(0.001)	(0.001)		(0.001)
Vantage score (Z, t-2)		-0.059***	-0.057***		-0.017***	-0.015***
		(0.000)	(0.000)		(0.001)	(0.001)
Sex $\times$ Ethnicity FE	Yes	Yes	Yes	Yes	Yes	Yes
Age + Age^2	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.008	0.041	0.042	0.014	0.015	0.018
Observations	821,392	821,392	821,392	169,376	169,376	169,376
Top20-Bottom20	0.046		0.022	0.040		0.035
AUC	0.584	0.720	0.721	0.653	0.670	0.685
E(Y)		0.079			0.038	

Panel B

	Delinquency					
	Thick file			Thin file		
	(1)	(2)	(3)	(4)	(5)	(6)
Risk-taking (Z)	0.018***		0.010***	0.013***		0.011***
	(0.001)		(0.001)	(0.001)		(0.001)
Patience (Z)	-0.006***		-0.003***	-0.001		-0.001
	(0.000)		(0.000)	(0.001)		(0.001)
Vantage score (Z, t-2)		-0.059***	-0.058***		-0.017***	-0.016***
		(0.000)	(0.000)		(0.001)	(0.001)
Sex $\times$ Ethnicity FE	Yes	Yes	Yes	Yes	Yes	Yes
Age + Age^2	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.007	0.041	0.042	0.012	0.015	0.017
Observations	821,392	821,392	821,392	169,376	169,376	169,376
Top20-Bottom20	0.038		0.021	0.033		0.030
AUC	0.579	0.720	0.721	0.646	0.670	0.682
E(Y)		0.079			0.038	

Figure D1: ROC Comparison

This plot shows Receiver Operating Characteristic (ROC) curve to measure the performance of the lagged Vantage score alone (blue) versus the predictions from a ML model combining facial features with lagged Vantage score (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-past-due in the 12 months immediately preceding t (the window $[t-12, t]$).

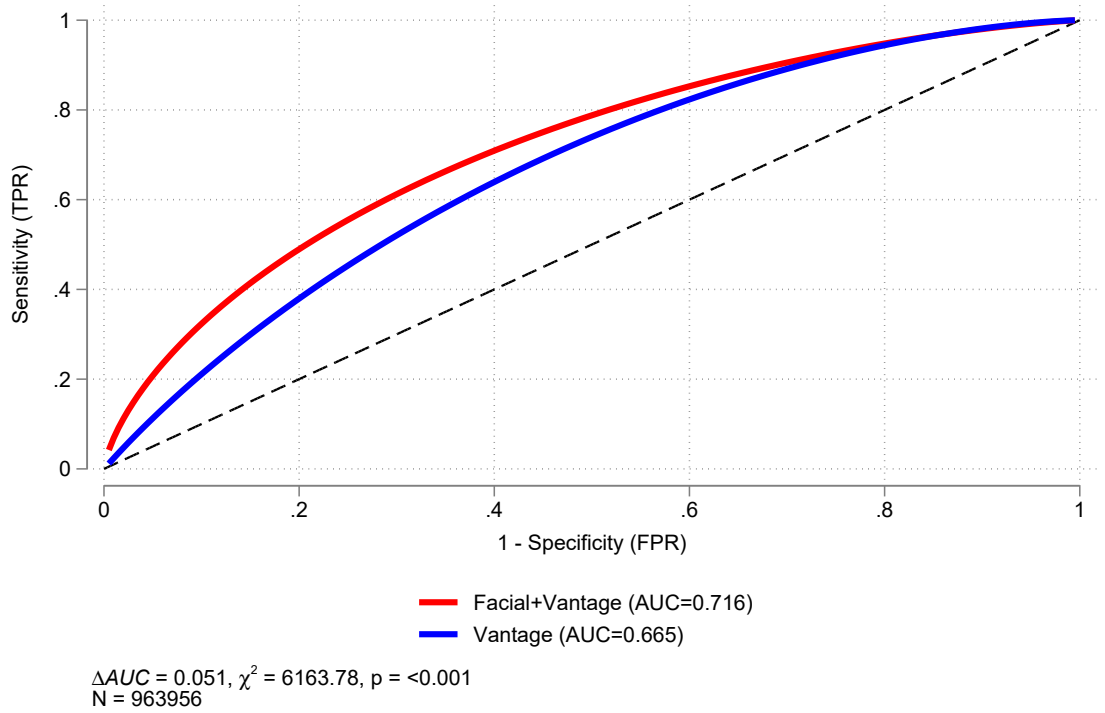


Figure D2: Number of Tradelines

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: a ML model using facial features only (green circles), lagged Vantage score only (blue squares), and a ML model combining facial features with lagged Vantage score (red diamonds). All three logistic regressions are 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-past-due in the 12 months immediately preceding t (the window $[t-12, t]$).

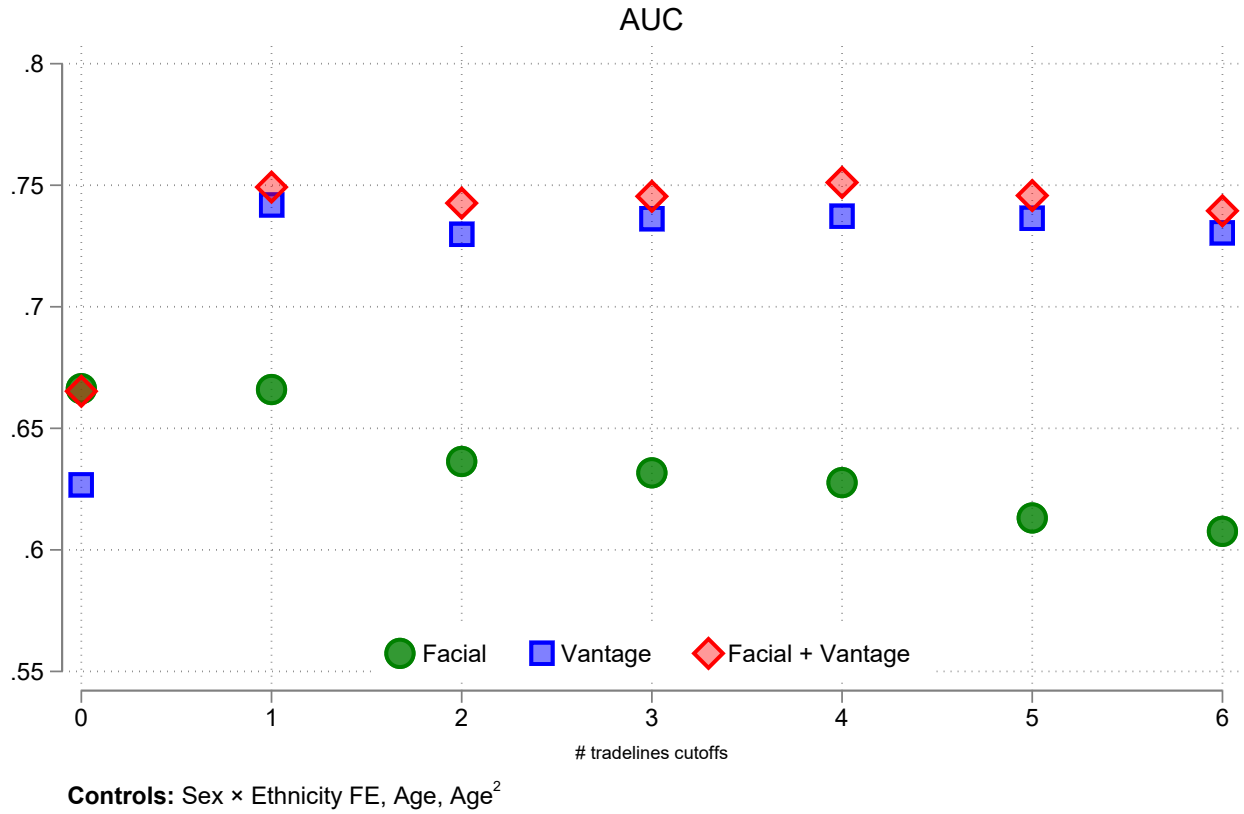


Figure D3: Borrower Age

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: a ML model using facial features only (green circles), lagged Vantage score only (blue squares), and a ML model combining facial features with lagged Vantage score (red diamonds). All three logistic regressions are 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-past-due in the 12 months immediately preceding t (the window $[t-12, t]$).

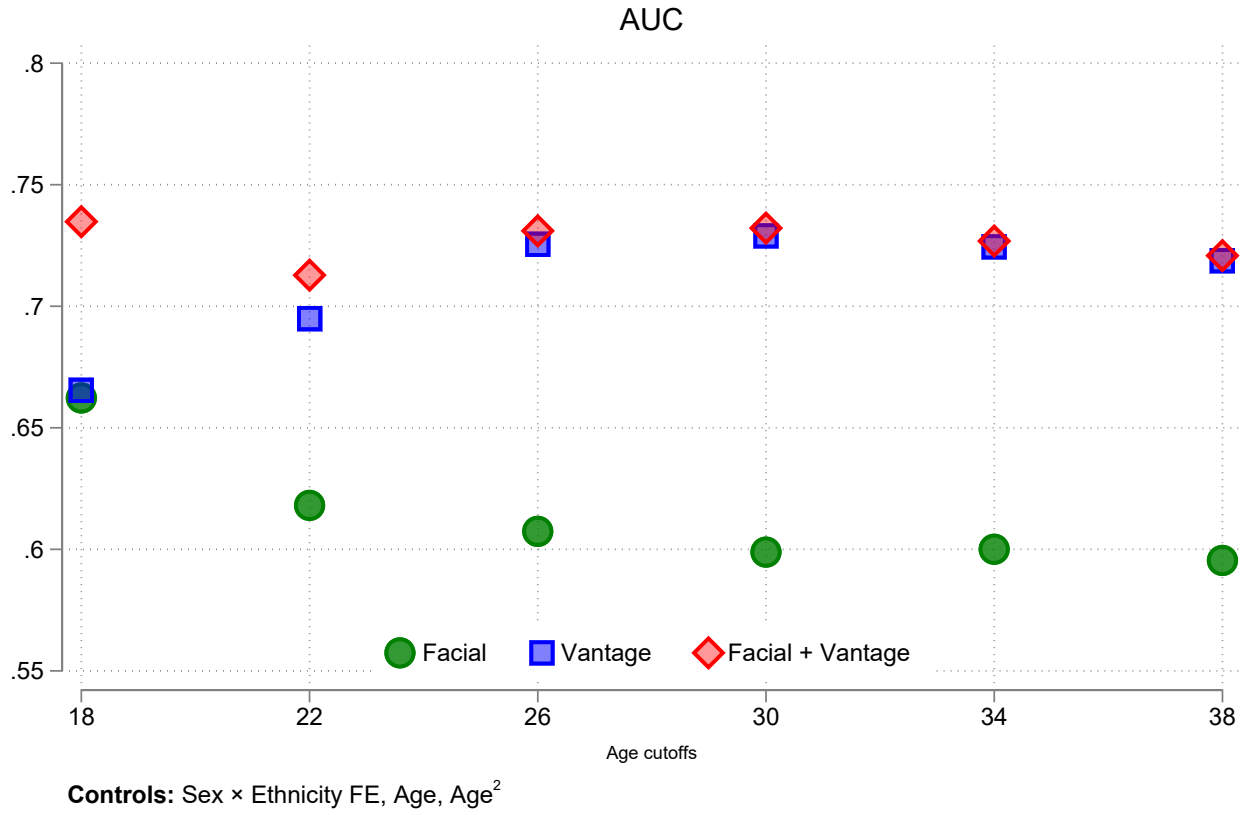


Figure D4: Borrower Income

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: a ML model using facial features only (green circles), lagged Vantage score only (blue squares), and a ML model combining facial features with lagged Vantage score (red diamonds). All three logistic regressions are 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-past-due in the 12 months immediately preceding t (the window $[t-12, t]$).

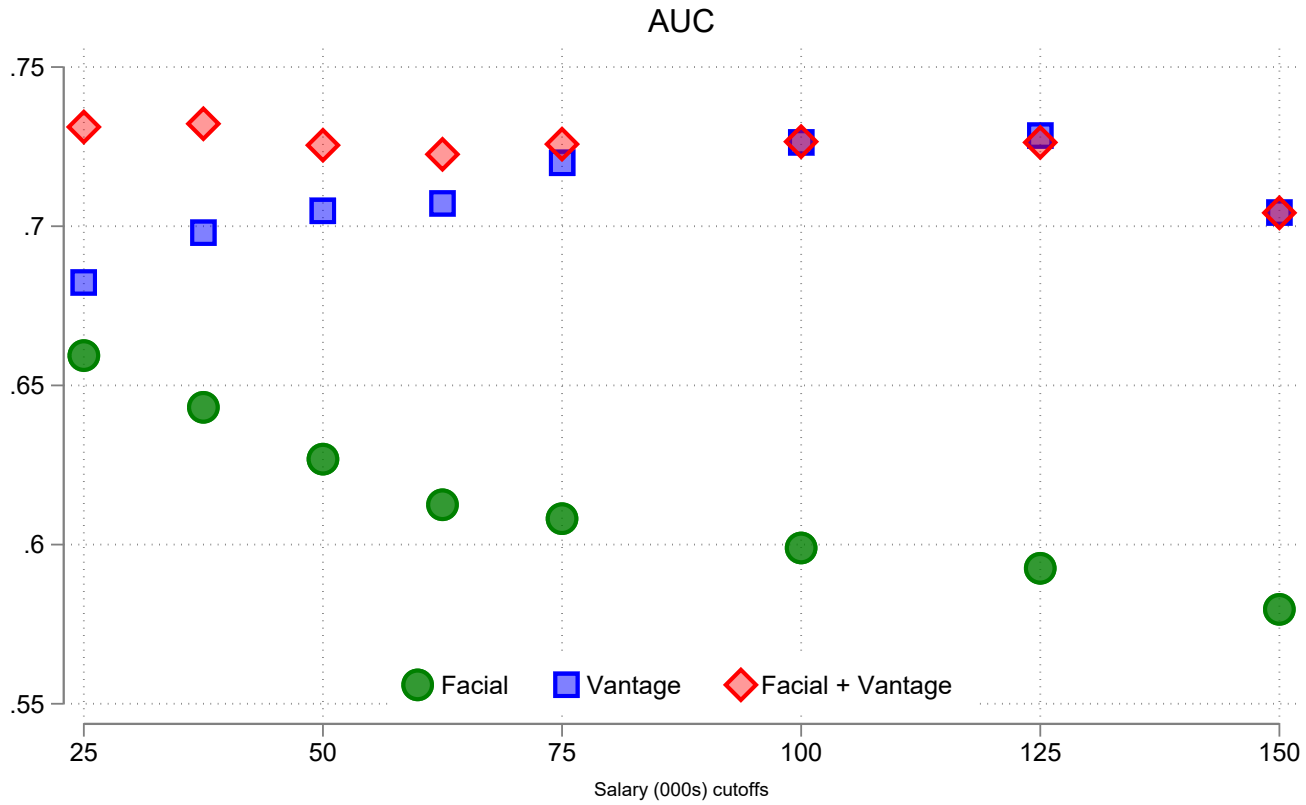
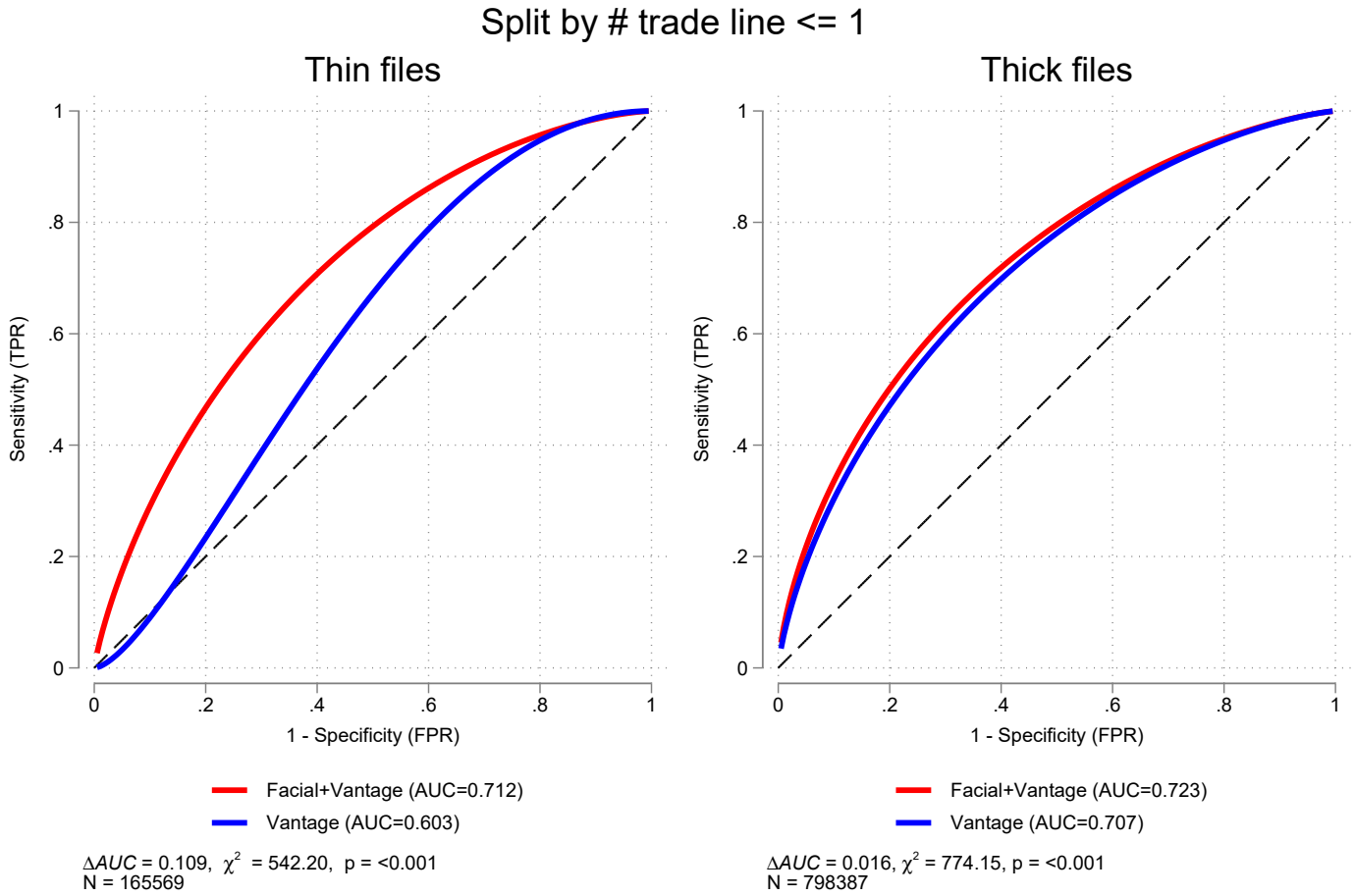


Figure D5: ROC Comparison for Thick versus Thin Files

This plot shows Receiver Operating Characteristic (ROC) curve to measure the performance of the lagged Vantage score alone (blue) versus the predictions from a ML model combining facial features with lagged Vantage score (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-past-due in the 12 months immediately preceding t (the window $[t-12, t]$).



E. Robustness: Analysis for a Predominant Borrower Group

Table E1: Delinquencies Prediction

This table regresses realized 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-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)		0.016***				0.016***	0.009***
		(0.000)				(0.000)	(0.000)
Log salary (Z)			-0.001***			0.011***	0.001**
			(0.000)			(0.000)	(0.000)
US school rank (Z)				0.009***		0.005***	0.004***
				(0.000)		(0.000)	(0.000)
No US school rank				0.005***		0.004***	-0.001
				(0.001)		(0.001)	(0.001)
Vantage score (Z, t-2)					-0.036***	-0.039***	-0.053***
					(0.000)	(0.000)	(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	318,297	318,297	318,297	318,297	318,297	318,297	318,297
Top20-Bottom20		0.045				0.043	0.024
AUC	0.557	0.580	0.511	0.541	0.676	0.694	0.719

Table E2: Thick versus Thin Files

This table regresses realized delinquencies on the facial delinquencies score from our ML predictions based on facial features alone and lagged Vantage score, separately for thin and thick-filed borrowers. *Delinquency* is defined as an indicator equal to 1 if the borrower experiences a 30+ days-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 $t-2$. 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***		0.009***	0.010***		0.008***
	(0.000)		(0.000)	(0.001)		(0.001)
Vantage score (Z, t-2)		-0.073***	-0.071***		-0.013***	-0.012***
		(0.001)	(0.001)		(0.001)	(0.001)
Age + Age^2	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.005	0.050	0.051	0.008	0.010	0.013
Observations	269,642	269,642	269,642	56,414	56,414	56,414
Top20-Bottom20	0.046		0.023	0.027		0.023
AUC	0.571	0.741	0.742	0.636	0.644	0.669
E(Y)		0.063			0.029	

Table E3: 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***			0.128***
		(0.002)			(0.002)
Log salary (Z)			0.292***		0.135***
			(0.002)		(0.002)
US school rank (Z)				-0.110***	-0.071***
				(0.002)	(0.002)
No US school rank				-0.049***	-0.093***
				(0.004)	(0.003)
Age + Age^2	Yes	No	No	No	Yes
R2	0.130	0.084	0.087	0.009	0.167
Observations	381,212	381,212	381,212	381,212	381,212
Top20-Bottom20		0.794			0.353

Figure E1: ROC Comparison

This plot shows Receiver Operating Characteristic (ROC) curve to measure the performance of the lagged Vantage score alone (blue) versus the predictions from a ML model combining facial features with lagged Vantage score (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-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 $t-24$.

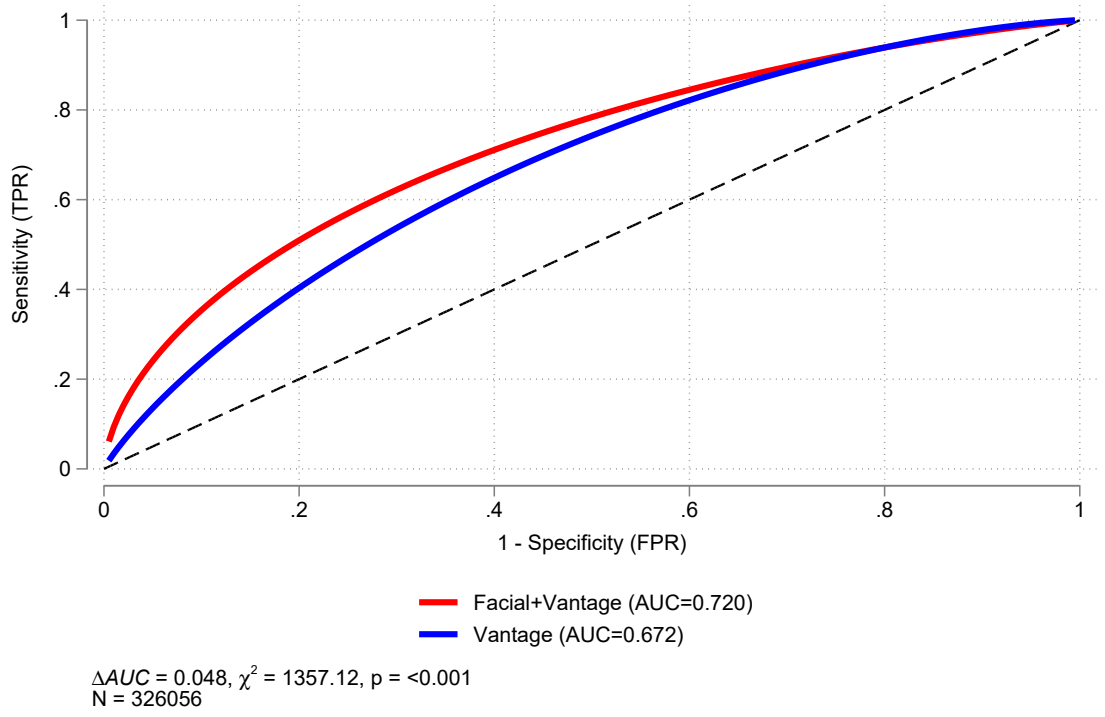


Figure E2: ROC Comparison for Thick versus Thin Files

This plot shows Receiver Operating Characteristic (ROC) curve to measure the performance of the lagged Vantage score alone (blue) versus the predictions from a ML model combining facial features with lagged Vantage score (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-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 $t-24$.

