Credit Scores and Committed Relationships*

Jane Dokko†
Federal Reserve Bank of Chicago

Geng Li‡
Federal Reserve Board

Jessica Hayes§
Rand Corporation

February 2020

Abstract

We document a large positive correlation in the credit scores of those entering committed relationships, even after controlling for similarities in socioeconomic and demographic characteristics. Moreover, we find that couples' average and match quality in credit scores, measured at the time of relationship formation, are highly predictive of subsequent separations, even beyond their ability to predict credit use and financial distress. Our results suggest that credit scores reveal an individual's relationship skill, and we present ancillary evidence to interpret this skill as one's trustworthiness and commitment to non-financial obligations.

Keywords: Credit scores, Committed relationships, Household finance, Trustworthiness

JEL codes: D14, G21, J12

*We thank Robert Putnam for his insights and encouragement. We also thank Stefania Albanesi, Meta Brown, Dan Hartley, Julie Hotchkiss, Julio Rotemberg, our colleagues at the Federal Reserve Board, and participants at numerous seminars and conferences for their helpful comments, and Waldo Ojeda and Lucas Nathe for excellent research assistance. The views presented in this paper are those of the authors and are not necessarily those of the Federal Reserve Bank of Chicago, the Federal Reserve Board, or its staff.

†Federal Reserve Bank of Chicago, Chicago, IL 60604. E-mail: jane.dokko@chi.frb.org

‡Federal Reserve Board, Washington, DC 20551. E-mail: Geng.Li@frb.gov.

§Rand Corporation, Santa Monica, CA 90401. E-mail: jessica.hayes.90@gmail.com.

Credit Scores and Committed Relationships

Abstract

We document a large positive correlation in the credit scores of those entering committed relationships, even after controlling for similarities in socioeconomic and demographic characteristics. Moreover, we find that couples' average and match quality in credit scores, measured at the time of relationship formation, are highly predictive of subsequent separations, even beyond their ability to predict credit use and financial distress. Our results suggest that credit scores reveal an individual's relationship skill, and we present ancillary evidence to interpret this skill as one's trustworthiness and commitment to non-financial obligations.

Keywords: Credit scores, Committed relationships, Household finance, Trustworthiness

JEL codes: D14, G21, J12

1 Introduction

With the rapid expansion in the availability and use of household credit over the past three decades, credit scores have become omnipresent in households' financial and non-financial decisions and opportunities. For example, credit scores are a key input into virtually all mortgage, credit card, and auto loan underwriting and thus affect households' access to credit, borrowing costs, as well as their ability to purchase durable goods and to smooth consumption over the lifecycle or against income fluctuations. Credit scores are also frequently used in other areas besides debt underwriting, such as auto insurance contracts, cell phone plans, and rental housing. Moreover, survey evidence suggests that up to 60 percent of employers run credit checks on potential employees as part of the hiring decision (Chen et al., 2013; Bos et al., 2015).

Motivated by the growing prominence of credit scores in everyday life, we explore their role in partner selection and relationship dissolution using a large, proprietary data set. Looking at marriages and long-term cohabitations, our results point to a quantitatively large and significant association between credit scores and the formation and dissolution of committed relationships. Specifically, we document three empirical facts. First, credit scores are positively correlated with the likelihood of forming a committed relationship. Second, individuals tend to select partners with similar credit scores as themselves, even after controlling for other similarities between partners. Third, in spite of this positive correlation, within-couple differences at the start of relationships exist and are highly predictive of subsequent separations. Their predictive power holds even after controlling for other important factors that can be directly affected by couples' credit scores, such as their use of credit and the occurrence of financial distress.

These results lead us to hypothesize that credit scores, in addition to measuring an individual's creditworthiness regarding the repayment of debt obligations, reveal information about an important relationship skill. We provide ancillary, corroborative evidence that this skill relates to an individual's trustworthiness and commitment to non-debt obligations.[1] Indeed, we

[1]Note that we are not the first to connect creditworthiness and character. In 1913, at a Congressional hearing, asked whether credit was "based primarily upon money or property," the banker J.P. Morgan replied "No, sir;

find that, similar to how credit scores predict the formation and dissolution of committed relationships, survey-based measures of trustworthiness also predict the outcomes of relationships. Moreover, we show that survey-based measures of trustworthiness are correlated with the average credit score of a geographic area (typically a county). We also find that when individuals have longer exposure to greater trustworthiness, they tend to have higher credit scores, even years after leaving those areas. Importantly, we note that we are not proposing that credit scores are equivalent to trustworthiness. We acknowledge that there are many factors unrelated to trustworthiness, such as labor market and health shocks, that can affect one's credit scores. In spite of this, we argue that, for certain borrowers, their credit scores reveal information about their underlying trustworthiness.

Our study contributes to a wide range of topics related to (i) how individuals sort into committed relationships, (ii) the economic inputs of successful and failed relationships, and (iii) factors beyond economic inputs that contribute to the durability of committed relationships. On the first topic, we identify a new characteristic by which individuals sort themselves into committed relationships, complementing a large literature that finds that race, educational attainment, earnings capacity, parental wealth, social caste, and physical appearance are also traits by which individuals sort (Weiss and Willis, 1997; Garfinkel et al., 2002; Blackwell and Lichter, 2004; Watson et al., 2004; Charles et al., 2013; Banerjee et al., 2013; Chiappori et al., 2012).

On the second topic, we find evidence that the mismatch in credit scores constrains household durable goods consumption, contributes to financial distress, and ultimately, undermines relationship durability. These results highlight the importance of credit markets for households and support the idea that consumption complementarities—as they manifest in the credit market—are the raison d'être of committed relationships (Avery et al., 2009; Chatterjee et al., 2009; Han et al., 2015; Stevenson and Wolfers, 2007; Lam, 1988). In addition, we find one dimension along which relationship outcomes are predictable—namely the initial match quality in credit

the first thing is character."

scores, supplementing the prevailing view that relationship dissolution is largely unpredictable and often owes to unfavorable and unpredictable outcomes that transpire during the relationship (Becker et al., 1977). With the emergence of new dating websites that allow individuals to reveal information about their creditworthiness, there has been heightened media attention on the importance of credit scores in relationship building and maintenance (Silver-Greenberg, 2012; Ettin, 2013). To the best of our knowledge, our paper is the first to provide systematic evidence that supports the assumptions behind this nascent trend of match-making.

On the third topic, we argue that credit scores are important for many economic choices of households outside of the credit market. We find that couples' credit score mismatch remains predictive of relationship dissolution after controlling for credit use and financial distress, which themselves may account for relationship outcomes. We conduct ancillary, corroborative analysis suggesting that credit scores may be correlated with relationship skills, such as one's trustworthiness, which also influences one's commitment to non-debt obligations. Indeed, because households are subject to fewer formal, contractual restraints and use more implicit contracts that can be difficult to enforce, relationship skills are all the more important in maintaining committed relationships and mitigating problems arising from incomplete information, incomplete contracts, and lack of enforceability (Arrow, 1972; Weiss, 1997; Guiso et al., 2004; Karlan, 2005). In particular, our results point to how relationship skills help overcome some of the problems arising within relationships due to incomplete information and incomplete contractability.

Of note, on a conceptual level, couples' income and differences therein are likely to have significant effects on relationship outcomes. However, our analysis does not observe individual income and instead uses an approximation to control for the potentially confounding effect of income. That said, we argue that the association between credit scores and relationship formation and dissolution is not merely picking up a residual income effect due to the potential correlation between credit scores and income. First, credit scoring algorithms deliberately omit income from their models, suggesting limited mechanical ability for income to predict credit

scores, conditional on the other variables in credit scoring algorithms. Indeed, using a random sample of households where income and credit scores are both contemporaneously observed, Beer et al. (2018) find only a modest correlation (about 0.3 to 0.4) between income and credit scores. The relatively low correlation is corroborated by many media reports of high-income individuals and celebrities who fail to pay their bills on time and eventually become to financially insolvent. Second, we find that the within-couple correlation in credit scores is substantially larger than that observed in partners' income (0.63 versus 0.38) estimated in household survey data, reassuring a less-than-perfect income-credit score correlation. That said, research on the relationship between credit scores and income has been somewhat inconclusive, and some analysis indicates a stronger correlation (see, for example, Albanesi et al. (2019)). More research and better data are needed to further establish the relationship between credit scores and income.

The remainder of the paper is organized as follows. Section 2 presents institutional background on credit scoring and introduces a simple conceptual framework for interpreting the role of credit scores in relationship formation and dissolution. Section 3 discusses the different sources of data used in this paper and introduces the algorithm we use to identify relationship formation. Sections 4 and 5 present the statistical and econometric framework and results, respectively, and Section 6 concludes and sets an agenda for future research. The Methodological Appendix provides details on and validation of the algorithms used to identify the formation and dissolution of committed relationships and highlights the methodological contributions of this paper.

2 Institutional Background and Conceptual Framework

2.1 Background on Credit Scoring

Credit scores evaluate the credit risk of potential borrowers and reflect a rank ordering that corresponds to a borrower's likelihood that s/he will miss payments and default on an account at some point in the near future (typically over the next two years). In constructing credit

scores, credit reporting agencies' algorithms consider an individual's debt payment history as the primary determinant but other factors, such as levels of indebtedness, age of the credit history, credit limit utilization, new inquiries, and public judgments (such as bankruptcy, foreclosure, tax liens, and garnishment) are also important contributing factors (Chatterjee et al., 2011).

Importantly, the agencies' credit score algorithms intentionally ignore information on gender, age, monthly income, assets, employment history, and occupation in estimating an individual's credit score (Beer et al., 2018). As mentioned earlier, previous research is inconclusive as to the extent to which income is correlated with credit scores. On the one hand, Beer et al. (2018) find a modest correlation of 0.3 to 0.4 in a random sample of consumers whose income and credit scores are both observed. Moreover, conditional on a parsimonious set of credit history variables, household income has essentially no additional explanatory power in predictive credit scores. On the other hand, Albanesi et al. (2019) argue for a strong correlation between the two and that, after age, income is the most important predictor of credit scores. Resolving this debate is not an aim of this study, and, instead, we provide an interpretation of our results acknowledging differing views about the determinants of credit scores.

Today, credit scores have wide applications in both credit and non-credit market contexts. In the former, nearly all banks and other large financial institutions use credit scores to underwrite and price loans to households. Generally, those with higher credit scores are deemed more creditworthy and likely to meet their debt repayment obligations. Individuals with higher credit scores typically have more abundant access to credit and face lower borrowing costs (all else equal). Non-credit market uses of credit scores include background checks in the labor, rental, and auto insurance market. Survey evidence suggests that up to 60 percent of employers, including the federal government, use credit checks in their hiring decisions (Chen et al., 2013). Auto insurance providers in nearly all states take credit information into account in estimate the risk of car accidents while many cell phone and cable companies use credit score information in contract-based plans.

2.2 Conceptual Framework

A key insight of the theoretical literature on committed relationships is that sorting occurs in partner selection and relationship formation across various models. For example, Lam (1988) focuses on consumption complementarities as the main gains of committed relationships and postulates that partners' incomes are positively correlated when positive matching in income allows for greater consumption gains. In contrast, other models, most notably the canonical marriage market model in Becker (1973), typically predict a negative correlation between partners' incomes as complementarities in the production of household output drive partner selection. In the framework of consumption complementarities, higher credit scores enable better access to the credit market and lower costs of financing. Other factors, including income, held constant, higher credit scores imply greater consumption capacity.

In addition, research has underscored the self-insurance motive against income shocks to smooth household consumption (Blundell et al., 2008). In this sense, higher credit scores imply greater capacity for self-insurance and potentially less volatile consumption within couples. The consumption-complementarity consideration in relationship formation would also lead to positive matching on credit scores.

Above and beyond their direct role in predicting debt repayment behavior and credit access, we further conjecture that credit scores may contain signals that reveal the underlying trustworthiness of an individual. We test this hypothesis by examining whether credit scores continue to predict relationship outcomes after credit use and realized financial distress are controlled for. Our hypothesis can be illustrated in the following conceptual framework. Broadly speaking, a person's default probability is affected by her willingness to repay debts, which we denote with ω, and her ability to repay, which we denote with η. Thus, one's credit score, as a (noisy) indicator of her default probability, can be represented as

$$score = f(\omega, \ \eta) \ + \ \mu, \tag{1}$$

where μ is a random error term. In this set-up, willingness to repay debts is, in turn, closely

related to an individual's underlying trustworthiness. We believe there is a basis for making this connection. Historically, lenders have long recognized a borrower's general trustworthiness and character as important factors influencing her debt payment history and default probability. We find that the credit reports garnered in the 1930s, in addition to debt repayment history and other financial data, collected information on borrowers' characteristics, reputation, habits, morals, and even illegal liquor traffic activities. For example, one of the credit reports prepared by the Retail Credit Company in 1934 included the following questions: "Does his record show he has been a steady and reliable man?" "Is his personal reputation as to character, honesty, and fair dealing good?" "Do you learn any illegal liquor traffic activities or domestic difficulties?"[2]

While modern credit reports are prohibited from collecting such information, credit scores may continue to reflect trustworthiness because they directly reflect past debt repayment history and defaults, which are themselves related to borrowers' commitment to debt obligations. Accordingly, in many types of transactions outside of the credit market, credit scores are used, among other purposes, as an indicator of trustworthiness. Accordingly, as discussed earlier, besides loan underwriting and pricing, credit scores are used extensively in the rental, labor, and auto insurance markets, though Dobbie et al. (forthcoming) argue that employers' use may have limited labor market consequences given the range of other screening tools used to assess job qualifications.

3 Data and Algorithm

3.1 The Consumer Credit Panel/Equifax Data

For individual-level data on credit scores and the formation and dissolution of committed relationships, we turn to the Federal Reserve Bank of New York's Consumer Credit Panel/Equifax (henceforth the CCP). Equifax is one of the three largest U.S. credit reporting agencies. These data are extensively used to analyze how households' liabilities evolve over time. Beginning in

[2]The images of these credit reports are available from the authors upon request. See also Lauer (2017) for a detailed history of credit reporting.

the first quarter of 1999, the data are a quarterly panel and track a 5 percent random sample (the "primary sample") of all U.S. adult consumers with a valid credit history.[3] Our sample ends in the second quarter of 2014 and covers a total of 62 quarters. A unique feature of the CCP is its large sample size, with the primary sample containing about 12 million consumers in a typical quarter, which is crucial for our research design. All personally identifiable information is removed by Equifax.

In addition to the primary sample, the CCP also follows consumers who live at the same address as a consumer from the primary sample. However, these non-primary sample consumers are followed only for the duration they share the same address. Once they stop living with primary sample consumers, non-primary consumers and their credit records are no longer observed. Their information is also not observed before they live with primary sample consumers. Roughly 30 million non-primary sample consumers live with someone in the primary sample in a typical quarter.

Variables in the CCP include: a credit score (known as the Equifax Risk Score); information on loan balances; the delinquency status of various types of debt; bankruptcy, foreclosure, and other derogatory flags; and various geographic identifiers down to the census block. Like most other credit scores, the Equifax Risk Score is designed to predict the likelihood of severe delinquency over the next 24 months and ranges from 280 to 850, with a higher score indicating higher credit quality.[4]

3.2 Identifying Relationship Formation

Taking advantage of the large sample size in our data (roughly 12 million consumers in a typical quarter), we design a novel algorithm to identify committed relationships, when they form, and when they dissolve. The general idea to identify committed relationships is to find pairs

[3]The randomness of the sample derives from including only individuals whose last two digits of their social security numbers belong to a pre-specified set of five numbers. The last four digits of the social security number are assigned sequentially to new applicants in chronological order as applications are processed and are thus as good as randomly assigned. Such randomness is a feature of, for example, Johnson et al. (2006) and Gross et al. (2014).

[4]For more information about the CCP, see Lee and van der Klaauw (2010).

of individuals who had not lived together but at some point start sharing the same address. The quarter in which they begin sharing the same address marks the "start" of the committed relationship.[5] Because not all people living together are in committed relationships, we apply a sequence of restrictions and perform several robustness checks to ensure that most of the consumer-pairs we identify are indeed couples.

We first apply our algorithm to only the individuals in the primary sample who we are able to follow both before and after relationship formation, as well as after relationship dissolution should that occur. Applying the steps of our algorithm leads to a sample of nearly 50,000 couples that form within the primary sample (that is, both partners belong to the primary sample). The algorithm is summarized as the following steps:

1. *In each quarter, Q, find the pairs of primary sample individuals who live at the same address in that quarter but have lived in difference addresses between Q-8 and Q-1.* This step identifies the start of a committed relationship for those couples who start living together in Q but have not lived together at the same address in the previous 8 quarters.[6]

2. *Keep only the pairs where the two individuals are the only people living at that address. The couple is removed if they also live with non-primary sample individuals as of Q.* This step restricts the identified couples to those who were the only individuals who live at that address. This is needed because some individuals who live in the same apartment building or dorm are recorded as having the same address in the Equifax data. In addition, this step help reduce the chances of roommates being counted as couples by removing the addresses shared by three or more roommates.

3. *Keep only the pairs where both individuals are between age 20 and 55 as of Q and the age difference between them is twelve years or less.* This step lets us focus on prime-age indi-

[5]Dettling and Hsu (2014) use a similar algorithm to find the adult children who moved back to parents' home in the same data.

[6]In one robustness check, we strengthen the restriction that couples do not share the same address 16 quarters prior to the quarter when we observe that they move to live together.

viduals and exclude the possibility of adult children moving in to live with their parents. Moreover, ninety-nine percent of couples observed in various nationally representative household survey data report age differences of less than 12 years.

4. *Keep only the pairs identified who stay living together from Q to at least $Q + 4$.* This step requires couples to live together for at least five quarters to restrict the sample to reasonably stable relationships to further sift out roommate and other non-committed relationships, many of which do not last longer than on year.

We also apply a modified version of this algorithm to all consumers in the primary and non-primary CCP samples to find over 2 million couples. Because we only observe non-primary sample individuals when they share the same address with primary sample individuals, we cannot apply the restriction of not sharing the same address between Q-8 and Q-1 on relationships involving non-primary sample individuals. Accordingly, we say that a primary and non-primary consumer form a committed relationship in quarter Q if Q is the first time that the non-primary consumer appears in the data. All other steps of the algorithm apply. We use this expanded set of couples for robustness analysis and extensions.

Importantly, we note that our algorithm cannot distinguish legally married couples from those in a stable, cohabiting relationship. However, this distinction is not critical for our analysis because we are interested in the implications of credit scores and the associated match quality in a general swathe of committed relationships, not just the couples who are legally married. Indeed, the marriage rate has experienced a steady decline since the 1990s, whereas cohabitation has become an acceptable living arrangement in society (Lundberg and Pollak, 2014; Wang and Parker, 2014). Moreover, many cohabiting relationships eventually evolve into marriages, further blurring the distinctions between the two types of committed relationships (Stevenson and Wolfers, 2007). Lastly, cohabiting couples also share many household economic and financial responsibilities in a way similar to married couples, making credit scores an important subject to study even in this context.

The Appendix describes the many exercises we undertake to validate this relationship formation algorithm using both administrative and survey data. In short, we: (1) match relationship formation rates found in other survey and administrative data; (2) observe similar demographic features of couples as in other data sets where marital and cohabiting status is directly observed; and (3) analyze placebo couples, which allows us to reject the possibility that the algorithm is identifying randomly-matched individuals. These validation exercises confirm that most of the relationships we have identified represent true committed relationships.

3.3 Algorithm for Relationship Dissolution

Our strategy of identifying the couples that separate follows a similar idea. Specifically, among the identified relationships, we consider a couple as having separated in quarter Q if the two individuals begin living at different addresses for at least five quarters starting from Q and are never observed to share the same address again. Note that only separations within the sample of identified couples are studied as the objective of our analysis is to relate the outcome of a relationship to the conditions prevailing at the start of the relationship. About 15 percent of these couples separate by the end of the relationship's second year, and 15 percent of the couples whose relationship survive the first two years separate by the end of the fourth year. The relatively high dissolution rates are consistent with the possibility that some of the households are non-marital cohabitations that, on average, have a higher rate of separation than marriages.

3.4 Other Data

One limitation of the Equifax CCP data is that the dataset does not provide any demographic and socioeconomic information besides age. We therefore augment the Equifax CCP data using the 2000 Decennial U.S. Census. Specifically, we use the block group-level estimates of demographic and socioeconomic characteristics as approximations for characteristics of the individuals who live in those block groups.

4 Econometric Framework

Our analysis aims to address three questions: do credit scores predict the formation of committed relationships? What is the match quality in couples' credit scores, above and beyond other between-partner similarities? Do couples' credit scores predict relationship outcomes (that is, are match quality and the average level predictive)? Three main components of our statistical or econometric analysis correspond to these questions. First, in the cross-section of single individuals, we analyze the odds of forming a committed relationship in a given year by estimating various specifications of the following model:

$$R_y^i = \alpha + \beta Score_y^i + \gamma Age_y^i + \theta Z^i + \omega S^i + \zeta Year_y + u_y^i. \tag{2}$$

Here, R_y^i is a binary variable for whether individual i, who is single in the fourth quarter of year y, forms a committed relationship, as defined in Section 3.2, by the fourth quarter of year $y+1$.[7] $Score_y$ is the level of the credit score or a vector of credit score bins for individual i, depending on the specification, in year y, when the individual remains single. Age_y is a cubic polynomial of the individual's age in year y. Z is a vector of proxies for individual socioeconomic characteristics that include the log of median income, share of college educated, and share of white residents of the census block group where individual i is observed to have resided the longest time prior to year y. In the Appendix, we include a detailed discussion of using Census block group-level estimates as proxies for individual socioeconomic and demographic characteristics. S is a vector of state or county fixed effect to control for regional variation in how people sort into committed relationships as well as geographic patterns in credit scores.[8] $Year$ is a vector of year fixed effects to control for macroeconomic conditions and aggregate trends in relationship formation and credit scores. In some specifications, we allow for interacted fixed effects of $S \times Year$ to account for regional variations in these trends. The parameter of interest is β, and our conjecture

[7] Such relationships include those with nonprimary-sample individuals.

[8] For instance, on average, credit scores tend to be lower in the Southeast where marriage markets may also operate differently, as evidenced by the younger average age at first marriage.

is $\beta > 0$ as singles with higher credit scores tend to have greater access to credit and better consumption smoothing capabilities, which in turn makes them more attractive to a larger set of partners in the relationship market.

For the second question, among the committed relationships we identify, we estimate match quality at the time of relationship formation and the extent to which we see positive sorting by credit scores. We also estimate the joint distribution of credit scores and compare the distribution of what would arise with random matching among relationship-eligible sample members. In the last part of answering our second question, we analyze how couples' match quality evolves over time and contrast such dynamics between couples who stay together and those who separate.

Third, we estimate the following model to study how *initial* average credit scores and match quality are associated with the odds of separating and to explore the potential channels:

$$\begin{aligned} D^i_{q_1,q_2} &= \alpha + \beta_a \overline{Score}^i_{q_0} + \beta_g Score\ Gap^i_{q_0} + \gamma Age^i_{q_1} + \gamma_g\ Age^{gap,i} + \theta_g Z^{gap,i} + \delta Div^i \\ &\quad + \nu CM^i_{q_1,q_2} + \omega S^i + \zeta Year_{q_1} + u^i_{q_1,q_2}. \end{aligned} \tag{3}$$

Here, $D^i_{q_1,q_2}$ is a binary variable that indicates whether the couple, i, conditional on remaining in the relationship as of quarter q_1, separates by quarter q_2. Specifically, denoting q_0 as the quarter of relationship formation, we examine the prospects of relationship dissolution within the three windows: $(q_1, q_2) \in \{(q_0, q_0 + 8), (q_0 + 8, q_0 + 16), (q_0 + 16, q_0 + 24)\}$.[9] $\overline{Score}_{q_0}$ is the average credit score of the two partners measured at the time the committed relationship forms, and $Score\ Gap_{q_0}$ is the credit score difference observed at that time.

Among couples comprising two primary sample individuals, the longitudinal structure of the Equifax data allows us to follow both partners before and after the formation of their relationship. This feature offers critical advantages to circumvent endogeneity concerns in estimating the model. In particular, as two partners often share financial responsibilities and decisions,

[9]These time windows correspond to the second year, the third and fourth year, and the fifth and six year into the relationship. The relationship identification algorithm requires that the couple cannot break up within the first year of the relationship.

along with influencing each other on their attitudes and approaches of managing finances, their credit scores tend to converge during the course of their relationship. However, the credit scores of couples who separate do not converge, a pattern that we will explore in detail in the next section. Thus, the credit score match quality observed during a relationship can be correlated with the prospects of the relationship itself. In contrast, using *initial* credit score match quality observed at the time of relationship formation and relating this measure to *subsequent* relationship dissolution largely circumvents this endogeneity concern.[10]

Age and Age^{gap} denote age polynomials of both partners and their age difference, respectively. In addition, we include Z^{gap}, a vector of gaps in partners' socioeconomic and demographic proxies (measured using statistics of the census block group each partner resided before entering the relationship), to help us understand how such differences contribute to relationship dissolution—an unsettled question in the literature. This inclusion also helps to control for factors that may confound the relationship between dissolution and credit scores. Lastly, we include the local divorce rate, Div, observed in the 2000 decennial Census to control for the potential social stigma of divorce or separation. For couple i, Div corresponds to the 2000 census block group-level divorce rate in the couple's neighborhood where the couple resides as of q_1 so that separations in our sample do not feed back into the control variable.

To study the channels by which initial credit score conditions of the couple influence future relationship outcomes, we include CM_{q_1,q_2}, a vector of credit market outcomes realized between q_1 and q_2, in the model. CM includes both measures of credit use, such as joint account ownership and new borrowing, and indicators of financial distress, such as bankruptcy, foreclosure, and other derogatory records. Indeed, mechanically credit scores are designed to predict such credit market experiences, which themselves are associated with relationship outcomes. Finding that $\overline{Score}$ and $Score\ Gap$ have additional bearing on relationship outcomes after controlling for credit market experiences, therefore, speaks to our hypothesis that credit scores may affect

[10] In one of the robustness analyses, we use the credit score gaps observed eight quarters prior to the relationship formation to further isolate the potential comovements.

relationship outcomes beyond credit market channels. Lastly, S_i and $Year_{q_1}$ are state (of the couple's residence prior to period (q_1, q_2) and calendar year fixed effects.

5 Results

This section presents the results of statistical and econometric analysis of our three questions and explores the channels by which initial credit score conditions may affect subsequent relationship outcomes. We also present ancillary evidence that corroborates our interpretation that credit scores reveal important relationship skills, such as trustworthiness and commitment to non-financial obligations on top of predicting the likelihood of defaulting on financial debt.

5.1 Main Results

5.1.1 Credit Scores and Relationship Formation

To begin, Table 1 presents the estimates of Equation (2). Here, we control for credit scores linearly and non-parametrically using 50-point credit score bins (consumers of lowest credit scores comprise the omitted group). Estimated odds ratios for our logit models suggest that individuals with higher credit scores are more likely to form committed relationships. We find a nearly monotonic trend with the relationship formation likelihood peaking for singles with credit score between 700 and 800. Standard errors are clustered at the individual level. We estimate that a one standard-deviation higher credit score corresponds to a 10 percent higher likelihood of entering a committed relationship in the following year, on average—a result that is evident across different specifications. In columns 1-2, only a polynomial of the individual's age is controlled for; in columns 3-4, year fixed effects and individual characteristics proxies are added as the previous literature documents that individuals sort into relationships by age, race, education, and earnings potential; state fixed effects are included in columns 5-6. To further control for the possible variation in relationship formation prospects associated with time and location, we include various combinations of state $\times$ year, county, and county $\times$ year fixed effects, respectively, in linear probability models (reported in columns 7-9). The estimated coefficients

are largely invariant to the inclusion of these fixed effects and are broadly consistent with the odds ratio estimates reported in columns 1-6. The consistency of the results presented in Table 1 against the extensive controls included suggests that other characteristics appear unlikely to drive the association between credit scores and relationship formation prospects and that credit scores potentially reflect a distinct individual trait by which partner selection occurs. Finally, we note that the estimated coefficients of individual characteristics proxies suggest that, consistent with previous studies (for example, Stevenson and Wolfers (2007)), singles with higher income, higher educational attainment, and who are white tend to have a greater chance of forming a relationship.

5.1.2 Credit Score Match Quality upon Relationship Formation

In the upper panel of Table 2, the average credit score of the couples that form in the primary Equifax sample is about 660 at the time of relationship formation, which is somewhat lower than the whole-sample average of 680 and likely reflects the younger-than-average age of those forming committed relationships. Regarding individuals' sorting into relationships, we find a large and positive within-couple credit score correlation of about about 0.6 at the time of relationship formation. The magnitude of this correlation is consistent with the notion that access to credit and the potential individual traits reflected in credit scores are important complements in determining couples' joint consumption and capability of weathering income shocks. Moreover, partners' credit score percentiles, rather than their levels, demonstrate the same pattern.[11]

One implication of such a strong, positive correlation is that the dispersion in credit scores at the couple level is similar to that at the individual level. To see this, note that the standard deviation of the credit score distribution across all individuals in our sample of identified couples is about 105. In a sample where individuals were randomly matched, the estimated standard deviation of the distribution of couples' average-credit score is about 75. However, in our sample of committed relationships, the standard deviation of couples' average credit score is

[11]The percentiles are calculated relative to the score distribution of all primary sample individuals in the quarter of relationship formation.

much higher at 92. Because a couple's access to credit and its prices (interest rates and fees) are often determined by the lower score of the two individuals, we further examine the dispersion of the minimum score between two partners. Its standard deviation is essentially identical to that across all individuals in our sample. Thus, the inequality in credit access across individuals is largely preserved across the couples involving these individuals. Put differently, access to credit and the ability to self-insure among low-credit score individuals do not appear to improve significantly after forming a relationship as their partners tend to have similarly limited access to the credit market.

The contingency table in the middle panel of Table 2 corroborates that credit scores are positively correlated within a couple and that the positive correlation is much larger than what we would expect to observe if individuals were randomly matched to their partners (presented in parentheses in the contingency table).

To put the within-couple credit score correlation upon relationship formation in context, it is helpful to compare this correlation with the correlations of other characteristics of the couple. For example, in the Panel Study of Income Dynamics, the correlations of being white, having college education, and individual labor income (of the first year of marriage) are 0.71, 0.41, and 0.38, respectively. Thus, the initial credit score correlation is about the same magnitude as the correlation in race but higher than that in education and income. This rank-ordering of correlations lends additional support that the behaviors represented by credit scores are likely an important and distinct individual trait that influences partner selection. Indeed, after projecting each spouse's credit score on an age polynomial and census block group-proxies of individual characteristics (race, educational attainment, and log median income), we continue to find that the within-couple correlation coefficient of the regression residuals is 0.50. The residual correlation is only slightly lower than the unconditional correlation of credit scores, suggesting that the within-couple credit score correlation is unlikely to be primarily driven by socioeconomic and demographic similarities.

Importantly, the matching in credit scores is not perfect and the estimated average within-

couple credit score differential is 69, shown at the bottom of the top panel. Putting this statistic in perspective, the estimated average differential between two randomly matched individuals is about 150. Hence, the within-couple credit score differential is slightly under one half of the difference between two randomly matched individuals. In addition, credit score differentials appear to vary with the average levels of credit scores of the couple.[12] Finally, because credit scores can fluctuate from quarter to quarter, it is important to rule out that the within-couple score differential estimated at the time of relationship formation represents a temporary change. We therefore examine the average score taken over across quarters $Q-3$ and Q to isolate the permanent component of the credit score. The results are qualitatively unchanged relative to those presented in Table 2.

5.1.3 Match Quality Dynamics

We now study how match quality in credit scores evolve after the formation of the relationship, which will shed light on the interactions between credit score match quality and the durability of the relationship itself. To begin, the upper panel of Figure 1 presents the dynamics of each partner's credit score during the first 16 quarters of the relationship (for relationships that last this long), with the blue (orange) curve tracking the average score of the individuals with higher (lower) scores at the time of relationship formation. As the figure shows, the difference in credit scores narrows appreciably over the first 16 quarters of the relationship, and the convergence is largely reflecting an increase in the credit score of the partners with lower initial scores.[13] The pattern is similar and more pronounced among couples who remain together longer. As shown in the lower panel, the average credit score gap among couples who remain together for at least 32 quarters (8 years) narrows to only 10 points by the eighth year in the relationship.

This convergence in credit scores likely reflects the shared financial responsibilities and mu-

[12]The percentile differential for the two inner quartiles of the average credit score distribution is on average about 20, comparing with an average 12 for the two outer quartiles, indicating that the match quality among middle-credit score couples is somewhat worse than that among low- and high-score couples.

[13]Interestingly, it is common for the within-couple ranking of credit scores to switch over time. After the first four years, over one-third of the couples see a reversal in the score ranking.

tually influencing behaviors and attitudes between partners rather than spurious or other mechanical factors (such as aging). Indeed, the top panel of Figure 2 shows no convergence of credit scores after couples separate. Similarly, among "placebo couples," we find no credit score convergence, noting that these randomly paired individuals are subject to the same age restrictions discussed earlier.

5.1.4 Initial Credit Score Conditions and Relationship Outcomes

In Table 3, we estimate the association between relationship dissolution and *initial* credit scores (Equation 3). Similar to Table 1, we use various specifications of controls and fixed effects to examine the robustness and consistency of our results. Columns 1-2 include no controls and no fixed effects, columns 3-4 include controls and year fixed effects, and columns 5-6 further include state fixed effects. Several salient features emerge in the results. First, couples with higher average credit scores are less likely to separate. Our odds ratio estimates indicate that a one standard deviation increase in couples' average credit scores (about 90 points) corresponds a more than 20 percent lower likelihood of separation during the second year of a relationship; and the associated reduction of separation likelihood is over 30 percent for longer horizons (the 3rd and 4th year and the 5th and 6th year).

Second, initial match quality, defined as the gap between two partners' credit scores, is highly predictive of relationship separation, even controlling for the average couple credit score. The odds ratios show that a one standard deviation increase in the score differentials (about 66 points) corresponds to an over 25 percent higher likelihood of separation during the second year of the relationship, over 20 percent during the third or fourth year, and over 10 percent during the fifth or sixth year. The estimates are largely unchanged when controlling for couples' average credit scores using more flexible 50-point bins (columns 2, 4, and 6). Moreover, estimating linear probability models that include state $\times$ year or county fixed effect yields largely similar results (columns 7-8). Finally, we note that the estimated coefficients of the control variables (not shown) are consistent with the notion that worse match quality on socioeconomic and

demographic characteristics tends to predict less stable relationships. For example, relationships with greater age and educational attainment differentials and relationships involving different races tend to have higher odds of separation. In addition, couples residing in communities with a higher (lagged) divorced population shares are more likely to separate. Overall, the results presented in Table 3 highlight one way in which relationship outcomes are predictiable, which complements the view that separations are largely shock-driven (Becker et al., 1977; Charles and Stephens, 2004).

The estimated association between initial credit score conditions and relationship outcomes are robust to an wide range of alternative econometric, sample selection, and relationship identification algorithm specifications, which we discuss in detail in the Appendix.

5.2 Exploring Potential Channels

We now explore the channels by which credit score match quality may affect relationship outcomes. We allow for a channel driven by economic outcomes and a channel for personal characteristics. The existing literature generally assigns stronger attribution to favorable and unfavorable economic outcomes—as opposed to partners' individual characteristics or their match quality—in explaining relationship dissolution (e.g. Charles and Stephens (2004)). Our results suggest that both channels play an important role in accounting for relationship outcomes. We first reaffirm that credit scores are indeed strong predictors of credit use and financial distress. We then show that observed credit use and financial distress predict relationship outcomes, giving support to the economic outcome channel. Finally, we show that initial credit score match quality continues to predict relationship dissolution even after controlling for the observed credit use and financial distress, which speaks to the personal characteristics channel.

5.2.1 Credit Usage and Financial Distress

In the upper panel of Table 4, we find that couples with worse match quality in credit scores are less likely to open a joint mortgage, auto, or credit card account in the first two or four years

of the relationship. For example, a one standard deviation increase in the initial credit score differential is associated with an over 35 percent lower likelihood of having a joint mortgage in the first two years (column 1). The estimated odds ratios are around 0.70 and 0.74, respectively, for opening joint auto loans and credit card accounts (columns 2-3). Looking over a four year horizon, the estimated odds ratios imply a somewhat weaker relationship between match quality and the use of joint accounts, likely reflecting the selection effects among the couples who stay together for at least four years. Similarly, as shown in the middle panel, greater initial credit score gaps also predict a lower likelihood of borrowing new mortgage and auto loans (regardless whether the debt were borrowed jointly).[14]

Second, we find that better matched couples are less likely to experience financial distress, especially in the first two years of the relationship. The results shown in the lower panel indicate that a one standard deviation increase in initial credit score gaps is associated with a 20, 6, and 17 percent higher likelihood of personal bankruptcy, home foreclosure, and derogatory records, respectively. Note that the coefficients in Table 4 are estimated controlling for county-level changes in unemployment and house prices. Adding these additional controls allows us to better isolate the effects of match quality from local economic conditions on credit market outcomes.

5.2.2 Two Potential Channels

Next, we argue that better matched couples are more capable to coordinate their financial and debt management strategies to support their joint consumption, weather income shocks, and manage financial risks, all potentially contributing to the stability of the relationship. We test this prediction by regressing the separation indicator on lagged indicators of opening any joint debt accounts, borrowing new mortgage or auto loans, and experiencing financial distress. Specifically, we correlate relationship dissolution during the third and fourth year with credit market events of the first two years (upper panel of Table 5) and dissolution during the fifth and

[14]Because of the revolving nature of credit cards, we focus on new debt borrowing only for mortgage and auto loans.

sixth year with credit market events of the first four years (lower panel). The logit regressions include the same controls as in column 5 of Table 3. The results indicate that first observed joint account ownership and new borrowing are strong predictors of lasting relationships. For example, among couples opening joint accounts in the first two years, their odds of separating during the subsequent two years are only a fraction (23 percent) of those who remain not having such joint accounts (column 1). To a lesser degree, borrowing new mortgage or auto debt also predicts lower chances of separation (column 3). In contrast, financial distress, on balance, predicts a higher chance of separation. For example, couples filing for personal bankruptcy in the first two years of the relationship are 54 percent more likely to separate over the next two years than other comparable couples (column 5).

Next, we test whether credit score match quality predicts relationship outcomes beyond its associations with credit use and financial distress. To do so, we estimate logit models similar to those reported in Table 3 but control for observed credit use and distress. The results, reported in columns 2, 4, and 6, consistently show that initial credit score gaps predict future separations independent of the observed credit usage and financial distress. Moreover, the estimated odds ratios associated with credit score match quality are smaller than those in Table 3, suggesting that some of its predictive power is indeed through predicting credit usage and financial distress. Finally, we estimate a "kitchen sink" style model that include all observed credit usage and financial distress as controls, and, as shown in column 7 credit score match quality continues to have sizeable, significant predictive power.

We note that some of the credit usage and financial distress indicators turn insignificant when credit score gaps are included, potentially due to their correlations. Accordingly, we first estimate linear probability models of separation indicators on credit usage and financial distress and then correlate the residuals of these regressions with credit score gaps. The results (not shown) are similar to those in Table 5 and continue pointing to match quality's predictive power beyond the observed credit use and financial distress.

5.3 Interpretation

One major factor that affects one's credit score is debt payment history, which can be affected by labor market stress and adverse credit market conditions. To understand the extent to which the initial credit score conditions of a couple reflect the two partners' previous (differential) labor and credit market experiences versus other idiosyncratic factors, we project the initial credit score of each partner on eight quarterly lagged unemployment, delinquency, and personal bankruptcy rates of the individual's county. We further construct the initial average credit scores and gaps predicted by labor market history and the corresponding residuals. We then estimate the model of relationship dissolution using, instead of the observed credit score averages and gaps, their predicted values, gaps, and residuals, and the results are reported in Table 6. We find that, through the duration of the relationship, residual levels and gaps of credit scores are consistently associated with subsequent separations, reassuring the role of the non-labor market component of credit scores in predicting relationship outcomes.

Accordingly, we interpret the residual predictiveness of credit scores as revealing information about an individual's relationship skill. Here, we provide additional discussion and analysis to interpret this relationship skill as one's trustworthiness and commitment to nonfinancial obligations. We recognize that the scope of our research is limited in what we can say about other relationship skills, such as sharing household operation obligations, resolving conflicts, or being able to compromise, but our results nonetheless bring important data, analysis, and insights to interpreting credit scores beyond their primary purpose and in the realm of committed relationships. Bricker and Li (2017) reaffirm the idea that credit scores reveal trustworthiness by inferring community average credit scores as an indicator for social trust and social capital. They show that this indicator is consistent with other indicators of social capital adopted in the literature and that residents of high-social capital neighborhood are more likely to invest in stocks (Guiso et al., 2004).

For this discussion, we turn to survey data in order to estimate the correlation between

credit scores and character. The 2000 Social Capital Community Survey (SCCS) is a survey of 375 to 1,500 adults in 41 communities that ask respondents trust-related questions, such as "whether most people can be trusted or you can't be too careful." These trust-related questions significantly overlap with those from the World Value Survey and the General Social Survey, and are the basis for our estimates of "character."

More specifically, we use the answers to these questions to infer the average trustworthiness of the communities covered by the SCCS. We make this interpretation because we conjecture that people who interact with more trustworthy counterparties in their neighborhoods tend to have a more trusting attitude. Moreover, Glaeser et al. (2000) find that answers to these questions are more indicative of one's own trustworthiness rather than how trusting of others he is. To be sure, whether answers to such survey questions measure trusting behavior or trustworthiness remains unsettled. For example, Fehr et al. (2003) challenges the results in Glaeser et al. (2000), while more recently, Sapienza et al. (2013) present evidence that shed light on why the SCCS/GSS/WVS questions may reveal trustworthiness. Using the individual-level question in the SCCS, we construct an index of trustworthiness that is equal to 100 times the share of survey respondents in a community who reply that "most people can be trusted."[15] Values of this index range from 36 to 66, and have a standard deviation of 8 across the 38 communities that are also covered in the CCP data. This index is highly correlated with the average credit scores of the corresponding areas in the CCP, with a Spearman rank correlation near 0.85, as shown in Figure 3. Moreover, this correlation persists after controlling for each survey community's level of income. As seen in column 1 of Table 7, when we project the average community credit scores onto the trustworthiness index, the estimated coefficient is highly significant, and the R-squared is above 50 percent. Adding the logarithm of the community's median income to the specification leaves the coefficient on the trustworthiness index little changed (column 2).

We then analyze whether previous exposure to trustworthiness is predictive of one's current

[15] In this calculation, we apply the SCCS weights to account for the stratified survey design.

credit score.[16] We construct a subsample of individuals in the CCP living in the communities surveyed by the SCCS for at least three years before moving out from these areas. We limit our analysis to those living in a community for at least three years to ensure sufficient time for any neighborhood effects to manifest. We then follow these individuals and estimate the correlation with contemporaneous credit scores. We find that previous exposure to neighborhood credit scores predicts contemporaneous credit scores. We acknowledge that this correlation is not a test for *causal* neighborhood effects as individuals likely sort into neighborhoods by their credit scores and other individual-level characteristics.

In our final set of observations, summarized in Table 8, we find that married individuals in the SCCS are more likely to report "most people can be trusted" than divorced or separated individuals (52 percent versus 42 percent). With our (and the literature's) interpretation of this variable, this difference may reveal that individuals remaining in relationships are more trustworthy. In line with this, there is a negative correlation (-0.37) between the share of separated or divorced respondents and the share of high-trustworthiness individuals across the SCCS-covered communities. These statistical links between relationship outcomes and trustworthiness provides additional support for interpreting credit scores beyond their primary purpose and suggests that they may reveal a relationship skill related to trustworthiness.

Lastly, in the bottom panel, we see that, among CCP couples where the partners live in an SCCS-covered community for at least three years prior to relationship formation, our trustworthiness index predicts the likelihood of separation. Specifically, a one-standard-deviation increase in the trustworthiness index is associated with a 6 percent lower likelihood of separating within the first six years of the relationship.[17] Moreover, once we control for the couples' average initial credit score, which we infer to be correlated with trustworthiness per the top panel in Table 8, we find that the coefficient on the trustworthiness index becomes small in magnitude and statistically insignificant in column 2. Under the null hypothesis that credit

[16] See Guiso et al. (2004); Brown et al. (2008) for evidence on enduring neighborhood effects on financial market participation.

[17] The standard errors are clustered at the SCCS community level.

scores extend beyond their primary purpose and reveal a relationship skill, we would expect a high degree of collinearity, as we do in column 2.

6 Conclusion

Using a large proprietary dataset of the credit records of U.S. consumers over fifteen years, we document that, conditional on observable socioeconomic and demographic characteristics, individuals in committed relationships have credit scores that are highly correlated with their partners' scores at the start of the relationship, consistent with consumption complementarities being an important motive in partner selection. We also find that the initial match quality in credit scores is highly predictive of relationship outcomes in that couples with larger score gaps at the beginning of their relationship are more likely to subsequently separate. One mechanism for this result is that poorly matched couples have a lower likelihood of using joint credit accounts, acquiring new credit, and avoiding financial distress. This result further supports the importance of consumption complementarities for relationship durability. However, we also find that the mismatch in credit scores is important for relationship outcomes beyond these channels, as evidenced by the residual predictive ability of credit scores conditional on credit usage and financial distress.

To interpret this last result, we analyze subjective, survey-based measures of trustworthiness and find support for interpreting the relationship skill revealed by credit scores as trustworthiness. First, we find that trustworthiness and credit scores are are positively and strongly associated at the neighborhood level, even controlling for (neighborhood) income. Moreover, an individual's neighborhood trustworthiness is predictive of his or her credit score, even long after leaving the neighborhood. Lastly, neighborhood trustworthiness predicts individual relationship outcomes, with a lower likelihood of separation among couples from more trustworthy areas. Contributing to a growing literature on the role of trust and social capital in supporting economic institutions and growth, our results present new evidence on how mismatch in trustworthiness within a household may affect its stability.

Our work potentially paves the way for two new areas of future research. First, essentially all of our knowledge on household finance derives from the household as the unit of analysis. Our study introduces an algorithm for identifying couples in credit bureau data and thus allows for additional analysis of how credit is allocated within a household and the implications of those choices. Second, future research is warranted to further explore the interpretation of credit scores as an objective, behavior-based, algorithm-derived measure of trustworthiness. Such research would push the boundaries of our understanding of trust and trustworthiness. After all, as Putnam (1995) famously wrote, "since trust is so central to the theory of social capital, it would be desirable to have strong behavioral indicators of trends in social trust and misanthropy. I have discovered no such behavioral measures."

References

Albanesi, Stefania, Giacomo DeGiorgi, and Jaromir Nosal, "Credit Growth and the Financial Crisis: A New Narrative," *Working Paper*, 2019.

Arrow, Kenneth, "Gifts and Exchanges," *Philosophy and Public Affairs*, 1972, *1* (4), 343–362.

Avery, Robert, Kenneth Brevoort, and Glenn Canner, "Credit Scoring and Its Effects on the Availability and Affordability of Credit," *Journal of Consumer Affairs*, 2009, *43* (3), 516–537.

Banerjee, Abhijit, Esther Duflo, Maitreesh Ghatak, and Jeanne Lafortune, "Marry for What? Caste and Mate Selection in Modern India," *American Economic Journal: Microeconomics*, 2013, *5* (2), 33–72.

Becker, Gary, "A Theory of Marriage: Part I," *Journal of Political Economy*, 1973, *81* (4), 813–846.

_ , Elizabeth Landes, and Robert Michael, "An Economic Analysis of Marital Instability," *Journal of Political Economy*, 1977, *85*, 1141–1187.

Beer, Rachael, Felicia Ionescu, and Geng Li, "Comparing Three Credit Scoring Models," *FEDS Notes 2018-05-21*, 2018.

Blackwell, Debra and Daniel Lichter, "Homogamy among Dating, Cohabiting, and Married Couples," *The Sociological Quarterly*, 2004, *45* (4), 719–737.

Blundell, Richard, Luigi Pistaferri, and Ian Preston, "Consumption Inequality and Partial Insurance," *American Economic Review*, 2008, *98* (5), 1887–1921.

Bos, Marieke, Andres Liberman, and Emily Breza, "The Labor Market Effects of Credit Information," *Working Paper*, 2015.

Bricker, Jesse and Geng Li, "Credit Scores, Social Capital, and Stock Market Participation," *FEDS Working Paper No. 2017-008*, 2017.

Brown, Jeffrey, Zoran Ivkovic, Paul Smith, and Scott Weisbenner, "Neighbors Matter: Causal Community Effects and Stock Market Participation," *Journal of Finance*, 2008, *43* (3), 1509–1531.

Charles, Kerwin and Melvin Stephens, "Job Displacement, Disability, and Divorce," *Journal of Labor Economics*, 2004, *22* (2), 489–522.

_ , Erik Hurst, and Alexandra Killewald, "Marital Sorting and Parental Wealth," *Demography*, 2013, *50* (1), 51–70.

Chatterjee, Satyajit, Dean Corbae, and José-Víctor Ríos-Roll, "Credit Scoring and the Competitive Pricing of Default Risk," *Working Paper*, 2009.

_ , _ , **and José-Víctor Ríos-Rull**, "A Theory of Credit Scoring and Competitive Pricing of Default Risk," *Unpublished manuscript*, 2011.

Chen, Daphne, Dean Corbae, and Andy Glover, "Can Employer Credit Checks Create Poverty Traps? Theory and Policy Evaluation," *Unpublished manuscript*, 2013.

Chiappori, Pierre Andre, Sonia Oreffice, and Climent Quintana-Domeque, "Fatter Attraction: Anthropometric and Socioeconomic Matching on the Marriage Market," *Journal of Political Economy*, 2012, *120* (4), 659–695.

Dettling, Lisa and Joanne Hsu, "Returning to the Nest: Debt and Parental Coresidence among Young Adults," *Federal Reserve Finance and Economics Discussion Series*, 2014.

Dobbie, Will, Paul Goldsmith-Pinkham, Neale Mahoney, and Jae Song, "Bad Credit, No Problem? Credit and Labor Market Consequences of Bad Credit Reports," *Journal of Finance*, forthcoming.

Ettin, Erika, "A Little Number Goes a Long Way," *Philadelphia Inquirer*, 2013.

Fehr, Ernst, Urs Fischbacher, Bernhard von Rosenbladt, Juergen Schupp, and Gert Wagner, "A Nation-Wide Laboratory. Examining Trust and Trustworthiness by Integrating Behavioral Experiments into Representative Surveys," *CESifo Working Paper 866*, 2003.

Garfinkel, Irwin, Dana Glei, and Sara McLanahan, "Assortative Mating among Unmarried Parents: Implications for Ability to Pay Child Support," *Journal of Population Economics*, 2002, *15* (3), 417–432.

Glaeser, Edward, David Laibson, Jose Scheinkman, and Christine Soutter, "Measuring Trust," *Quarterly Journal of Economics*, 2000, *115* (3), 811–846.

Gross, Tal, Matthew Notowidigdo, and Jialan Wang, "Liquidity Constraints and Consumer Bankruptcy: Evidence from Tax Rebates," *Review of Economic Statistics*, 2014.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, "The Role of Social Capital in Financial Development," *American Economic Review*, 2004, *94* (3), 526–556.

Han, Song, Benjamin Keys, and Geng Li, "Information, Contract Design, and Unsecured Credit Supply: Evidence from Credit Card Mailings," *Working Paper*, 2015.

Johnson, David, Jonathan Parker, and Nicholas Souleles, "Household Expenditure and the Income Tax Rebates of 2001," *American Economic Review*, 2006, *95* (5), 1589–1610.

Karlan, Dean, "Using Experimental Economics to Measure Social Capital and Predict Financial Decisions," *American Economic Review*, 2005, *83* (5), 1688–1699.

Lam, David, "Marriage Markets and Assortative Mating with Household Public Goods," *Journal of Human Resources*, 1988, *23* (4), 462–487.

Lauer, Josh, *Creditworthy: A History of Consumer Surveillance and Financial Identity in America*, Columbia University Press, 2017.

Lee, Donghoon and Wilbert van der Klaauw, "An Introduction to the FRBNY Consumer Credit Panel," *Federal Reserve Bank of New York Staff Reports*, 2010.

Lundberg, Shelly and Robert Pollak, "Cohabitation and the Uneven Retreat from Marriage in the U.S., 1950-2010," in Leah Platt Boustan, Carola Frydman, and Robert A. Margo, eds., *Human Capital in History*, National Bureau of Economic Research, 2014.

Putnam, Robert, "Bowling Alone," *Journal of Democracy*, 1995, pp. 65–78.

Sapienza, Paola, Anna Toldra, and Luigi Zingales, "Understanding Trust," *The Economic Journal*, 2013, *123*, 1313–32.

Silver-Greenberg, Jessica, "Perfect 10? Never Mind That. Ask Her for Her Credit Score.," *New York Times*, 2012.

Stevenson, Betsey and Justin Wolfers, "Marriage and Divorce: Changes and Their Driving Forces," *Journal of Economic Perspectives*, 2007, *21* (2), 27–52.

Wang, Wendy and Kim Parker, "Record Share of Americans Have Never Married," *Pew Research Report*, September 2014.

Watson, David, Eva Klohnen, Alex Casillas, Ericka Nus Simms, Jeffrey Haig, and Diane Berry, "Match Makers and Deal Breakers: Analyses of Assortative Mating in Newlywed Couples," *Journal of Personality*, 2004, *72* (5).

Weiss, Yoram, "The Formation and Dissolution of Families: Why Marry? Who Marries Whom? And What Happens Upon Divorce?," in M.R. Rosenzweig and O. Stark, eds., *Handbook of Population and Family Economics*, Amsterdam: Elsevier Science B.V., 1997.

_ **and Robert Willis**, "Match Quality, New Information, and Marital Dissolution," *Journal of Labor Economics*, 1997, *15* (1), S293–S329.

Table 1: Credit Score Levels and Spousal Relationship Formation

	Logit						OLS		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Credit score	1.147***		1.094***		1.094***		1.010***	1.010***	1.010***
	(0.001)		(0.001)		(0.001)		(0.000)	(0.000)	(0.000)
Score 450-499		1.011*		1.032***		1.032***			
		(0.006)		(0.006)		(0.006)			
Score 500-499		0.993		1.028***		1.026***			
		(0.005)		(0.005)		(0.005)			
Score 550-599		1.059***		1.086***		1.084***			
		(0.005)		(0.006)		(0.006)			
Score 600-649		1.147***		1.137***		1.134***			
		(0.006)		(0.006)		(0.006)			
Score 650-699		1.240***		1.183***		1.182***			
		(0.006)		(0.006)		(0.006)			
Score 700-749		1.435***		1.331***		1.330***			
		(0.007)		(0.007)		(0.007)			
Score 750-799		1.457***		1.310***		1.309***			
		(0.007)		(0.007)		(0.007)			
Score 800 +		1.302***		1.195***		1.193***			
		(0.008)		(0.007)		(0.007)			
Log(median income)			1.191***	1.192***	1.195***	1.197***	1.019***	1.018***	1.018***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Share college			0.952***	0.952***	0.952***	0.951***	0.995***	0.995***	0.995***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Share white			1.028***	1.027***	1.020***	1.019***	1.002***	1.001***	1.001***
			(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Control for									
Age polynomial	yes	yes	yes	yes	yes	yes	yes	yes	yes
Yearly FE	no	no	yes	yes	yes	yes	no	yes	no
State FE	no	no	no	no	yes	yes	no	no	no
Year × state FE	no	no	no	no	no	no	yes	no	no
County FE	no	no	no	no	no	no	no	yes	no
Year × county FE	no	no	no	no	no	no	no	no	yes
Pseudo R^2 and R^2	0.027	0.028	0.055	0.055	0.056	0.056	0.043	0.043	0.046
N	12,527,381	12,527,381	12,527,381	12,527,381	12,527,381	12,527,381	12,527,381	12,527,381	12,527,381

Note. The table reports odds ratios estimated from logistic regressions in columns 1-6 and linear regression coefficients in columns 7-9. Standard errors are clustered at the individual level and reported in parentheses. Credit scores, log of median income, share of college educated, and share of white residents are normalized by respective sample standard deviations. * denotes the estimate is statistically significant at the 90-percent level, and *** denotes the estimate is statistically significant at the 99-percent level. Columns 1, 3, 5 and columns 7-9 use the normalized levels of credit scores and columns 2, 4, and 6 uses dummies of credit score bins. Consumers with credit scores lower than 450 are the omitted group. Control variables of log median income, share of college educated, and share of white residents are census block group level statistics from the 2000 U.S. Census.

Table 2: Matching Quality at the Time of Household Formation

	Credit scores	
	Score levels	Score percentiles
	(1)	(2)
Mean credit score	657	41
Credit score standard deviations		
Cross-individual	104	26
Cross-couple (mean)	92	24
Cross-couple (minimum)	105	24
Within-couple correlation	0.59	0.63
Mean within-couple score differentials	69	17

	Contingency table: observed versus randomly matched (percent)		
	Prime	Near prime	Subprime
Prime	22.67 (11.73)	8.28 (10.30)	3.74 (12.66)
Near prime	7.58 (9.75)	12.56 (8.04)	8.68 (10.52)
Subprime	3.56 (12.34)	8.85 (10.83)	24.07 (13.32)
Column total	33.82	29.69	36.50

	Socioeconomic characteristics		
	CCP-identified	PSID-observed	CCP-placebo
	(1)	(2)	(3)
% White correlation	0.63	0.71	0.01
% College degree Correlation	0.49	0.41	-0.00
Median Income Correlation	0.41	0.38	0.02

Note: The upper panel presents statistics estimated for the identified couples that involve two primary-sample individuals. Score percentiles are calculated with respect to the population credit score distribution of the quarter of relationship formation. The middle panel presents the shares of couples by each spouse's credit score bin. Individuals with credit scores above 720 are labelled as prime, 620-719 as near prime, and below 620 as subprime. The numbers not in parentheses are observed shares, and the numbers in parentheses are the shares with individuals with different credit scores are matched randomly. In the lower panel, column 1 presents the statistics of the identified relationships that involve two primary sample individuals. Column 2 presents the statistics of the couples aged between 20 and 55 who are observed in the PSID data. Column 3 presents the results of the pseudo couples who are randomly matched in the CCP data subject to the same age and age differential restrictions. Statistics in columns 1 and 3 of the lower panel are estimated using the merged census block group level statistics in the 2000 U.S. Census, and those in column 2 are estimated using the survey-observed characteristics.

Table 3: Initial Credit Score Conditions and Subsequent Relationship Dissolutions

	Logit						OLS	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	The second year of relationship							
Initial score gap	1.343***	1.314***	1.288***	1.276***	1.285***	1.273***	1.037***	1.035***
	(0.016)	(0.016)	(0.017)	(0.017)	(0.017)	(0.017)	(0.002)	(0.002)
Initial score average	0.714***		0.783***		0.777***		0.974***	0.972***
	(0.010)		(0.012)		(0.012)		(0.002)	(0.002)
Initial score average bins	No	Yes	No	Yes	No	Yes	No	No
R2	0.0411	0.0436	0.105	0.105	0.107	0.107	0.109	0.136
N	43,889	43,889	40,323	40,323	40,125	40,125	41,676	41,233
	The third and fourth year of relationship							
Initial score gap	1.291***	1.275***	1.241***	1.235***	1.236***	1.231***	1.031***	1.031***
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.002)	(0.002)
Initial score average	0.621***		0.659***		0.652***		0.950***	0.947***
	(0.010)		(0.012)		(0.012)		(0.002)	(0.002)
Initial score average bins	No	Yes	No	Yes	No	Yes	No	No
R2	0.057	0.058	0.103	0.103	0.107	0.106	0.107	0.146
N	32,157	32,157	29,335	29,335	29,191	29,191	29,184	28,654
Controls								
Age polynomials	no	no	yes	yes	yes	yes	yes	yes
Age differentials	no	no	yes	yes	yes	yes	yes	yes
Current community char.	no	no	yes	yes	yes	yes	yes	yes
Community div. pop. share	no	no	yes	yes	yes	yes	yes	yes
Orig. community char. diff.	no	no	yes	yes	yes	yes	yes	yes
Yearly FE	no	no	yes	yes	yes	yes	no	yes
State FE	no	no	no	no	yes	yes	no	no
Yearly×state FE	no	no	no	no	no	no	yes	no
County FE	no	no	no	no	no	no	no	yes

Continued on next page

Table 3 – *Continued*

	Logit						OLS	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	The fifth and sixth year of relationship							
Initial score gap	1.166***	1.142***	1.131***	1.112***	1.125***	1.107***	1.011***	1.010***
	(0.025)	(0.025)	(0.026)	(0.025)	(0.026)	(0.026)	(0.002)	(0.002)
Initial score average	0.663***		0.686***		0.677***		0.968***	0.965***
	(0.016)		(0.017)		(0.018)		(0.002)	(0.002)
Initial score average bins	No	Yes	No	Yes	No	Yes	No	No
R2	0.034	0.038	0.069	0.072	0.075	0.078	0.069	0.126
N	24,052	24,052	20,620	20,620	20,526	20,526	20,518	19,914
Controls								
Age polynomials	no	no	yes	yes	yes	yes	yes	yes
Age differentials	no	no	yes	yes	yes	yes	yes	yes
Current community char.	no	no	yes	yes	yes	yes	yes	yes
Community div. pop. share	no	no	yes	yes	yes	yes	yes	yes
Orig. community char. diff.	no	no	yes	yes	yes	yes	yes	yes
Yearly FE	no	no	yes	yes	yes	yes	no	yes
State FE	no	no	no	no	yes	yes	no	no
Yearly×state FE	no	no	no	no	no	no	yes	no
County FE	no	no	no	no	no	no	no	yes

Note. The table reports odds ratio estimates in columns 1-6 and regression coefficients in columns 7-8. Standard errors are reported in parentheses. *** denotes the estimate is statistically significant at the 99-percent level. Odds ratios are calculated for a one-standard-deviation change of the respective variables. For logistic regression, report pseudo R-squared. The estimations reported in columns 2-7 control for a third-order age polynomial of each spouse and the age differential. These estimations also control for differentials in socioeconomic characteristics, using differences of U.S. Census statistics of race, education, and income between the census block groups where each spouse lived prior to forming the relationship as proxies. These estimations also control for the share of divorced population where the couple lived as of the end of the first, second, and fourth year of their relationship, respectively, for the three sets of results.

Table 4: Initial Credit Score Differentials and Subsequent Credit Usage and Financial Distresses

	First two years			First four years		
	(1)	(2)	(3)	(4)	(5)	(6)
-2ex			Open new joint loan accounts			
	Mortgage	Auto loan	Credit card	Mortgage	Auto loan	Credit card
$\frac{Initial\ diff}{100}$	0.644***	0.699***	0.737***	0.703***	0.833***	0.851 ***
	(0.022)	(0.022)	(0.038)	(0.023)	(0.025)	(0.042)
Pseudo R2	0.121	0.052	0.078	0.129	0.047	0.067
N	20,426	22,953	22,201	13,134	15,588	14,610
			Borrow new debt			
	Mortgage	Auto loan	Credit card	Mortgage	Auto loan	Credit card
$\frac{Initial\ diff}{100}$	0.938**	0.930***		0.925**	0.954*	
	(0.026)	(0.023)		(0.028)	(0.027)	
Pseudo R2	0.099	0.037		0.110	0.034	
N	13,217	12,182		8,492	8,697	
			New financial distress			
	Bankruptcy	Foreclosure	Derogatory records	Bankruptcy	Foreclosure	Derogatory records
$\frac{Initial\ diff}{100}$	1.202***	1.058	1.168***	1.013	1.035	1.023
	(0.044)	(0.057)	(0.023)	(0.040)	(0.057)	(0.026)
Pseudo R2	0.157	0.134	0.185	0.150	0.146	0.172
N	23,535	25,742	27,175	16,839	18,434	19,188

Note. The table reports odds ratio estimates. Standard errors are reported in parentheses. Odds ratios are calculated for a one-standard-deviation change of the respective variables. * denotes the estimate is statistically significant at the 90 percent level, ** denotes the estimate is statistically significant at the 95 percent level, and *** denotes the estimate is statistically significant at the 99 percent level. The estimations control for the array of bins of couple-average credit score at the time of relationship formation. The estimations control for a third-order age polynomial of each spouse and the couple age differential; These estimations also control for differentials in socioeconomic characteristics, using differences of U.S. Census statistics of race, education, and income between the census block groups where each spouse lived prior to forming the relationship as proxies. Finally, the estimations control for yearly and state fixed effects.

Table 5: The Role Played by Use of Credit and Financial Distress on Separations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Separated during the third or fourth year of relationship						
$\frac{Initial\ diff}{100}$		1.197***		1.227***		1.226***	1.193***
		(0.019)		(0.019)		(0.019)	(0.019)
Use of credit indicators							
Joint accounts	0.210***	0.259***					0.262***
	(0.014)	(0.019)					(0.018)
Mortgage			0.661***	0.720***			0.934
			(0.038)	(0.043)			(0.058)
Auto loans			0.863***	0.918			1.047
			(0.045)	(0.049)			(0.057)
Financial distress indicators							
Bankruptcy filing					1.543***	1.055	1.021
					(0.151)	(0.108)	(0.105)
Foreclosure					1.071	1.093	1.090
					(0.147)	(0.151)	(0.151)
Derog. records					1.817***	1.222***	1.198***
					(0.087)	(0.063)	(0.062)
Pseudo R2	0.099	0.127	0.073	0.108	0.077	0.107	0.127
N	30,228	29,191	30,228	29,191	30,228	29,191	29,191
	Separated during the fifth or sixth year of relationship						
$\frac{Initial\ diff}{100}$		1.091***		1.102***		1.107***	1.090***
		(0.025)		(0.025)		(0.026)	(0.025)
Use of credit indicators							
Joint accounts	0.421***	0.486***					0.512***
	(0.029)	(0.034)					(0.037)
Mortgage			0.581***	0.597***			0.715***
			(0.045)	(0.047)			(0.058)
Auto loans			1.083	1.112			1.206***
			(0.070)	(0.074)			(0.081)
Financial distress indicators							
Bankruptcy filing					1.397***	1.130	1.071
					(0.156)	(0.129)	(0.123)
Foreclosure					1.267	1.276*	1.255
					(0.187)	(0.188)	(0.186)
Derog. records					1.678***	1.202**	1.182**
					(0.118)	(0.089)	(0.088)
Pseudo R2	0.065	0.088	0.056	0.082	0.057	0.079	0.090
	21,026	20,526	21,016	20,526	21,026	20,526	20,526

Note. The table reports odds ratio estimates. Standard errors are reported in parentheses. Odds ratios are calculated for a one-standard-deviation change of the respective variables. *** denotes the estimate is statistically significant at the 99-percent level. Odds ratios are calculated for a one-standard-deviation change of credit score differentials and average levels at the time of relationship formation. All estimations control initial average credit score level bins as in table 3. The estimations control for a third-order age polynomial of each spouse and the couple age differential; These estimations also control for differentials in socioeconomic characteristics, using differences of U.S. Census statistics of race, education, and income between the census block groups where each spouse lived prior to forming the relationship as proxies. These estimations also control for the share of divorced population where the couple lived as of the end of the second and fourth year of their relationship, respectively. Finally, the estimations control for yearly and state fixed effects.

Table 6: Predicted and Residual Score Levels and Gaps

	The second year				The third and fourth year				The fifth and sixth year			
Predicted score level	1.015				0.937*				0.990			
	(0.026)				(0.027)				(0.045)			
Residual score level		0.735***				0.625***				0.677***		
		(0.011)				(0.011)				(0.017)		
Predicted score gap			1.074***				1.063***				0.993	
			(0.015)				(0.019)				(0.029)	
Residual score gap				1.255***				1.222***				1.147***
				(0.017)				(0.020)				(0.030)
Observed score levels			0.728***	0.767***			0.616***	0.643***			0.665***	0.687***
			(0.011)	(0.012)			(0.011)	(0.012)			(0.018)	(0.030)
N	42,091	39,563	40,145	39,393	30,213	28,749	29,177	28,656	20,997	20,213	20,499	20,163

Note. The table reports odds ratio estimates. Standard errors are reported in parentheses. Odds ratios are calculated for a one-standard-deviation change of the respective variables. Each spouse's credit score is projected on eight lagged quarterly unemployment rate, personal bankruptcy rate, and delinquency rate of the county that individual lived before forming the relationship, a third-order age polynomial, and yearly dummies. Credit score are decomposed into a predicted component and a residual, using which the couple's predict and residual credit score averages and gaps are calculated. Each regressor of Table 6 is normalized by the sample standard deviation. Odds ratios are evaluated at one-unit change of the regressors. * and *** denotes the estimate is statistically significant at the 90-percent and 99-percent level, respectively. The estimations control for a third-order age polynomial of each spouse and the couple age differential; These estimations also control for differentials in socioeconomic characteristics, using differences of U.S. Census statistics of race, education, and income between the census block groups where each spouse lived prior to forming the relationship as proxies. These estimations also control for the share of divorced population where the couple lived as of the end of the first, second, and fourth year of their relationship, respectively. Finally, the estimations control for yearly and state fixed effects.

Table 7: Survey Based Trustworthiness Index and Credit Scores

	Contemporary correlations		Long-term influences
	Community average credit score		Individual credit score
	(1)	(2)	(3)
Trustworthiness Index	1.57***	1.42***	
	(0.25)	(0.21)	
Trustworthiness Index			0.61***
(community lived 3 years ago)			(0.03)
Log(median income)		0.36***	0.42***
		(0.09)	(0.01)
R-squared	0.51	0.65	0.008
N	38	38	340,303

Note. The trustworthiness index is defined as 100 times the share of respondents who replied that "most people can be trusted" to the 2000 Social Capital Community Survey (SCCS) for each community surveyed, using the SCCS weights. Median income is the 2000 U.S. Census statistics of the corresponding community. Results reported in column 3 concern the individuals who lived in the communities covered by the survey for at least three years before moving out of these areas. The individuals' credit scores were taken three years after the move.

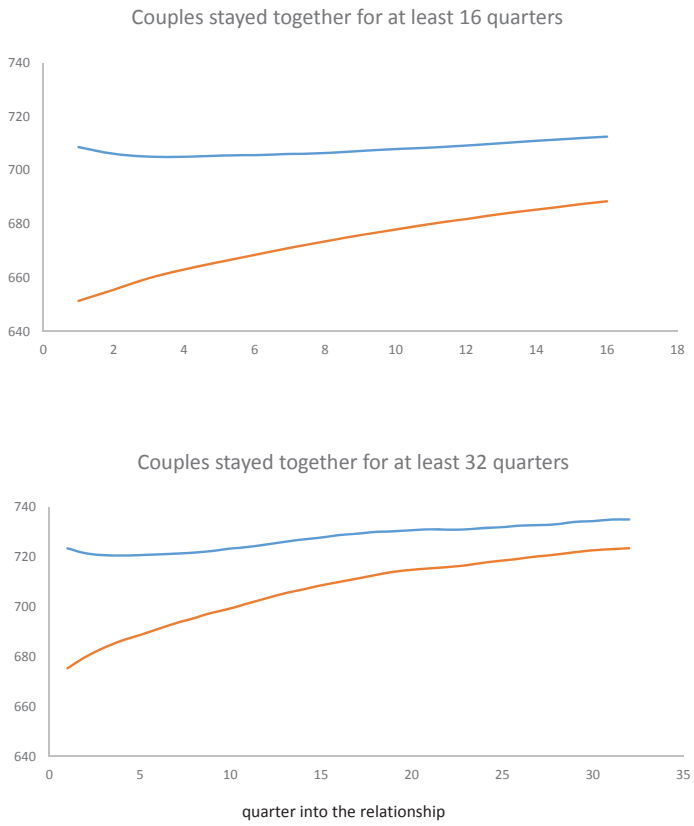
Table 8: Estimates of Relationships and Self-Reported Trustworthiness

SCCS respondents analysis		
Shares of individuals that have high trust levels	Married	Separated or divorced
	52.0%	42.3%
Correlations between		
Shares of the separated and divorced and shares of high trust levels		-0.37***

CCP couples analysis	(1)	(2)
Effects on separations within six years		
Trustworthiness index	0.941*	0.966
	(0.381)	(0.423)
$\frac{Initialscore}{100}$		0.599***
		(0.020)
Controlling for		
Age polynomials	Yes	Yes
Age diff.	Yes	Yes
Char. of community of rel. formation	Yes	Yes
Yearly FE	Yes	Yes
N	112,429	112,429

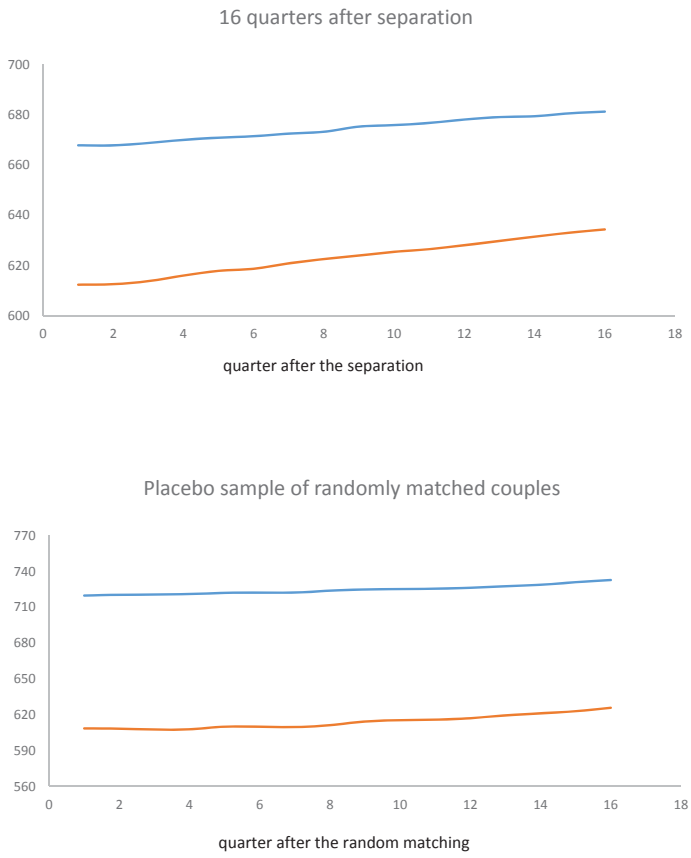
Note. Individuals of high trust levels refer to the respondents who replied that "most people can be trusted" to the 2000 Social Capital Community Survey (SCCS). The trustworthiness index is defined as 100 times the share of respondents who replied that "most people can be trusted" to the 2000 Social Capital Community Survey (SCCS) for each community surveyed, using the SCCS weights. The table reports estimated odds ratios associated with a one standard deviation increase of the independent variable. Standard errors are for the estimated coefficients and are clustered at the SCCS community level. * denotes the estimate is statistically significant at the 90 percent level, and *** denotes the estimate is statistically significant at the 99 percent level.

Figure 1: Dynamics of Credit Score Gaps of Lasting Relationships



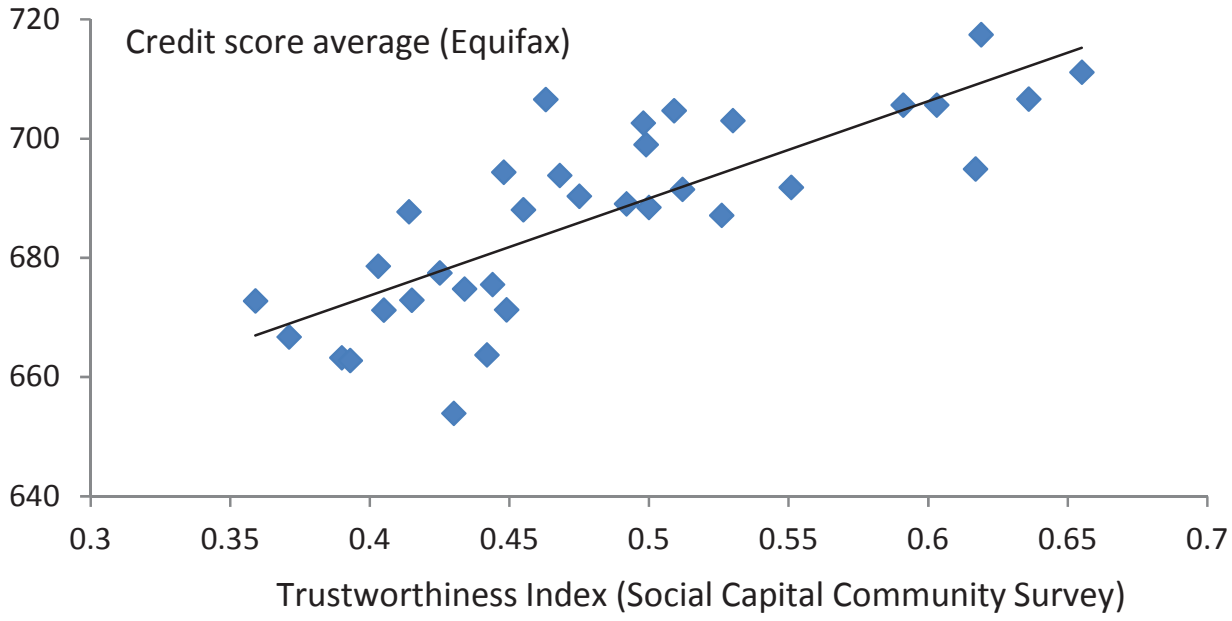
Source: Authors' calculation using the FRBNY Consumer Credit Panel / Equifax Data.
Note: The blue curve shows the average credit score of the individuals who had the higher score of a couple as of the time of relationship formation. The orange curve shows the average credit score of the individuals who had the lower score of a couple as of the time of relationship formation. The series are estimated using the couples where both individuals are in the primary sample.

Figure 2: Dynamics of Credit Score Gaps of Separated and Placebo Relationships



Source: Authors' calculation using the FRBNY Consumer Credit Panel / Equifax Data.
Note: The blue curve shows the average credit score of the individuals who had the higher score of a couple as of the time of relationship formation. The orange curve shows the average credit score of the individuals who had the lower score of a couple as of the time of relationship formation. The series are estimated using the couples where both individuals are in the primary sample. The placebo sample is constructed by randomly matching individuals of the primary sample, respecting the age and age differential restrictions imposed on the identified couples.

Figure 3: Credit Scores and Trustworthiness



Source: Authors' calculation using the FRBNY Consumer Credit Panel / Equifax Data and the Social Capital Community Survey.
Note: The figure plots the estimated trustworthiness index—the weighted share of respondents who replied that "most people can be trusted" to the 2000 Social Capital Community Survey of each surveyed community (horizontal axis) against the average credit score for the same community (vertical axis).

Appendix

A Validating the Algorithm

We conduct a sequence of validation exercises to reassure us that most of the couples identified by our algorithm represent committed relationships (that is, marriages or long-term cohabitations). We first compare the relationship formation rates in our sample with marriage rates in administrative and survey data. Then we examine the similarities in the demographic characteristics between partners of the committed couples we identified and compare them to couples in household survey data, where marital and cohabiting relationships are directly observed. In short, we find the relationship formation rates estimated using our couple-identification strategy compare favorably to the marriage rates estimated using the Panel Study of Income Dynamics (PSID) data and the related administrative statistics.[18] Also, the inferred socioeconomic and demographic characteristics between the spouses identified are correlated in a similar way as those of the spouses observed in household surveys, but such correlations are not detected in the placebo sample of randomly matched individuals.

A.1 Comparison of Relationship Formation Rates

Table A1 presents the relationship formation rates estimated using our algorithm. Nearly 50,000 couples form within the primary sample during the sample period, implying an annual relationship formation rate of 0.108 percent (column 1)—or 540 couples per one million individuals.[19] Because the baseline algorithm identifies only the couples formed among individuals of the primary sample, we are able to observe only a small fraction of the relationships formed as both spouses must be in the primary sample individuals. To see this, let Ω represent the population, let ω be a 5-percent random subset of Ω, and P_Ω be the national relationship formation rate. Individual $i \in \omega$ then has a probability P_Ω of forming a relationship with an individual $j \in \Omega$ in a given year. However, the probability of forming a relationship with $j \in \omega$, P_ω, should follow

$$P_\omega \approx 0.05 \times P_\Omega \tag{4}$$

Applying this adjustment gives the implied population relationship formation rate of 2.2 percent (shown in column 2). The marriage rate among individuals aged between 20 and 55 estimated using the Vital Statistics and the U.S. Census data is around 1.5 percent. Our estimated relationship formation rate appears a bit higher than these estimates, suggesting that our inclusion of cohabiting relationships seems to outweigh the exclusion of legitimate couples due to the selection criteria of our algorithm. When we apply the algorithm to include the non-primary sample, we obtain a relationship formation rate of 2.3 percent (column 3), which is in line with the adjusted primary sample formation rate. This broad consistency reassures us of the validity of our algorithm.

[18] The PSID is a nationally representative longitudinal survey of U.S. households. For more information about the PSID, see http://psidonline.isr.umich.edu.

[19] The relationship formation rate is calculated as the number of couples divided by the number of individuals aged between 20 and 55.

As seen in the subsequent rows of Table A1, the relationship formation rate declines with age, with the rate among individuals aged between 20 and 35 being above 2.6 percent while the rate among individuals aged between 46 and 55 below 1.5 percent. Although we do not observe marriage rates by age in administrative data, the decline in relationship formation rate is broadly consistent with the estimates of marriage rate derived using household survey data. For example, the marriage rates estimated using the PSID data are about 1.5 percentage points higher among the subsample aged between 20 and 35 than for the subsample between 45 and 55, a difference similar to that in our data. Also, the pattern of how the relationship formation rate varies across age groups is very similar between couples identified in the primary sample alone and when also using the non-primary sample.

A.2 Demographic Correlations: Do They Look Like Couples?

We now examine whether each member of the couples in the sample demonstrate similarities regarding their demographic and socioeconomic characteristics that are consistent with what we observe in the household survey data. Such information in the CCP data is very limited, partly due to federal laws prohibiting the collection and use of information on race, ethnicity, national origin, sex, and marital status in loan underwriting and in calculating credit scores. Indeed, because the only demographic information contained in the data is the year of birth, we supplement the CCP data with census block group level statistics from the 2000 U.S. Census. A census block group is a collection of multiple census blocks and typically has a population of approximately 1,000 people. Using block-group level averages, we approximate individual level demographic and socioeconomic characteristics, such as race, education, and income. For example, we use the share of adults in a block group who are white or have college degree to approximate individual race and educational attainment, and the census block group median income to approximate individual income.

As shown in Table A2, the average age of the individuals at the time of relationship formation is about 37 years, the average age differential is 3.6 years, and the within-couple age correlation is 0.85. These statistics are consistent with estimates from nationally representative surveys in which marital status is observed. For example, in the PSID, the average age at marriage, shown in column 2, is slightly younger at about 34 years. However, the age differentials and within-couple age correlation are remarkably similar between the CCP and the PSID estimates, reported in column 3. These estimates are also in line with previous estimates (Watson et al., 2004).

Using average block group level characteristics from the US Census, we show the correlation in partners' demographics in the bottom half of Table A2.[20] The approximated within-couple racial correlation is above 0.6, the college degree correlation is slightly below 0.5, and the income correlation is 0.35. Notably, the magnitudes of the estimated correlations are broadly consistent with previous studies, as is the rank ordering of their size (see, for example, respectively, Weiss and Willis (1997); Garfinkel et al. (2002); Blackwell and Lichter (2004)). Furthermore, the racial and income correlation coefficients are similar to those estimated using the 1999-2009 PSID

[20] We use the census block group of the sample member just prior to relationship formation if there is just one address or the characteristics of the block group where the individual lived for the longest duration for those with multiple addresses.

sample of newly married couples. While the PSID within-couple college degree correlation appears to be somewhat lower than the CCP estimates, the latter are very similar to the correlation estimated using other nationally representative household surveys, such as the Survey of Consumer Finances (correlation coefficient = 0.52) and the Consumer Expenditure Survey (correlation coefficient = 0.47).[21] Finally, we construct a placebo sample of randomly matched individuals in the primary sample that adhere to the age and age-difference restrictions imposed on the identified couples. As shown in column 3, we do not observe any appreciable correlations regarding race, educational attainment, and income within the pseudo-couples.[22]

A.3 Additional Motivation and Evidence for Census Block Group-Level Proxies

In Table A3, we show how using Census block group-level proxies for demographic characteristics – rather than individual-level measures – can lead to appropriate inferences about the magnitude of β in Equation 2 and β_a and β_g in Equation 3. Here, we turn to a PSID of households with annual labor earnings greater than $5,000 and estimate simple Mincer-style wage regressions. In column 1, we use demographic characteristics (race and education) that vary at the individual level. For illustrative purposes, suppose that we are interested in how the earnings of whites differ from non-whites but do not have individual-level data on education. This is an analogous situation to ours in estimating Equations 2 and 3 using the Equifax/CCP data, where we do not observe individual-level characteristics but want to address how their omission is likely to confound the interpretation of β, β_a, and β_g. Column 2 shows that omitting these variables from the regression leads to a larger estimate of the coefficient on the white indicator variable (0.335 versus 0.258), likely reflecting that white and educational attainment are positively correlated. In column 3, we control for education using the Census block-group proxies for education and the estimated coefficient of interest is much closer to the case where we observe all variables at the individual level (0.216 versus 0.258). In column 4, we use much cruder proxies for education that vary at the county level and find that, at this higher level of aggregation, using the proxies as controls is about the same as omitting education from the regression altogether and leads to a larger estimated coefficient than in column 1 (0.324 versus 0.258). This empirical exercise helps motivate the relevance of using Census block group-level proxies and provides suggestive evidence that such control variables may reasonably control for confounding individual-level factors that would otherwise bias the interpretation of our coefficients of interest.

B Robustness

Here, we provide a series of analyses to further support our conclusion that the initial match quality in the credit scores of those in committed relationships is predictive of subsequent separations. To begin with, recognizing the possibility that the pairs identified by our algorithm

[21] Indeed, the PSID data demonstrate a persistent decline in the within-couple college degree correlation over time, making the PSID statistics somewhat lower than those derived from other national surveys.

[22] The similarities regarding average age, age difference and age correlation are mechanical because we restrict the placebo couples to meet the same age and age difference restrictions.

may not be true, committed relationships, we are particularly concerned that the pairs with especially large differences in their credit scores are not true couples but are driving the results. Noting that estimates from the Current Population Survey find that less than 5 percent of two-person households who satisfy our restrictions are not in committed relationships, we discuss the results of a trimming exercise where we drop 5 percent of the observations with the largest initial score differences.[23] The results of this exercise show that the estimates in Table 3 are essentially unchanged.

The remaining results from additional robustness checks are reported in Table A4 and discussed below.

Using Credit Score Percentiles

First, we consider whether using credit score values in the match quality estimate may be affecting the interpretation of our results. Credit scores are developed to reflect a rank ordering of individuals' credit risk and so the distance between the two partners' credit scores may be measured by differences in their percentile ranking rather than by differences in the values of the credit scores. As shown in columns 1 and 2 in the upper panel of Table A4, the level and match quality effects of credit scores remain pronounced when credit score percentiles, rather than levels, are used. For example, all else equal, a one standard deviation increase of initial percentile differences corresponds to a 30 and 16 percent higher likelihood of separation by the end of the fourth and sixth years of the relationship, respectively.[24] This similarity should not surprise us as we obtain similar patterns in the degree of assortative matching and measuring within-couple differences in credit scores when percentiles are used (see Table 2).

Inferring the Timing of Relationship Formation

Two methodological choices affect how we link pre-relationship credit score conditions to relationship outcomes: assigning the quarter in which a couple begins living together as the marker of the start of a committed relationship and the length of the period during which a couple cannot be observed to live together before they start their relationship. On the first choice, we assume in the baseline analysis that credit score differentials measured in the quarter in which a couple first begins living together appropriately measure the match quality in credit scores. It may be that partners' financial behaviors influence one another prior to when they begin living together. Moreover, it could also be that one partner in a couple does not immediately change his or her mailing address precisely at the same time that the committed relationship begins. To address this concern, instead of the relationship formation quarter Q, we use each partner's credit score eight quarters before Q. These results, shown in columns 3 and 4 of Table A4, are qualitatively similar to the baseline estimates of separation odds ratios.

On the second choice, it may be that some couples use separate addresses even after forming a committed relationship, such as would be the case where an employee who is sent to work at a different location on a long-term basis than his or her home city. In this case, our baseline algorithm may mistakenly treat their moving back to the same location as a new relationship. To address this concern, we adjust the algorithm to limit the two partners sharing the same address

[23] These results are available upon request.

[24] The results of this and other robustness test specifications regarding relationship outcomes as of the end of the second year are also qualitatively unchanged relative to the baseline specification.

to 16, rather than 8, quarters before Q, the quarter in which the relationship forms. These results, shown in columns 5 and 6, again are qualitatively similar to the baseline estimates.[25]

Including Couples Involving Nonprimary Sample Individuals

We now examine whether the baseline results are sensitive to adding couples that include non-primary sample individuals. Allowing such couples being included in the analysis substantially enlarged the sample of couples.[26] However, because the first time we observe the non-primary sample spouse is when that individual began living at the same address with a primary sample individual, we do not observe the census block group the non-primary sample individuals lived prior to forming the relationship. As a result, we cannot control for Z^{gap}, the approximated socioeconomic and demographic gaps, as we did in the baseline analysis. The results are reported in columns 7 and 8 in the lower panel of the table, and the implied odds ratio estimates are very similar to those in the baseline analysis.

Focusing on Couples with Joint Credit Accounts

One possible concern with our sample is that our algorithm of identifying committed relationships may actually include pairs of individuals who are not in a cohabiting or marital relationship. To rule out this possibility, we focus our attention on couples sharing joint debt accounts as non-committed couples are unlikely to jointly apply for credit cards, mortgages, or auto loans and share the financial obligations of repayment. We show, in columns 9 and 10, that for such couples the initial credit score match quality remains predictive for future separations.

[25]The sample sizes are smaller in this specification because we do not observe all individuals of the baseline sample couples up to 16 quarters prior to relationship formation.

[26]However, because other variables are more likely to be left blank for non-primary sample individuals, the numbers of observations used in the estimations are smaller than those reported in Table A1.

Table A1: Annual Rates of Committed Relationship Formation and Separation

	Formation rates		
	Primary sample only		Whole sample
	Unadjusted (1)	Adjusted (2) = (1) × 20	(3)
Age 20-55	0.108%	2.16%	2.26%
Age 20-35	0.131%	2.62%	2.93%
Age 36-45	0.116%	2.32%	2.35%
Age 46-55	0.068%	1.36%	1.27%
Number of couples identified	49,363		2,070,117

	Separation rates	
	Primary sample only	Whole sample
During the 2nd year	15.1%	19.2%
During the 3rd and 4th year	14.9%	21.1%
During the 5th and 6th year	8.1%	12.4%

Note: For each age range, formation rates in column 1 are estimated as the ratios between the number of primary-sample individuals involved in committed relationships, as identified in the algorithm discussed in Section 3, and the total number of primary sample individuals. Column 2 presents the adjusted relationship formation rates, accounting for the fact that relationships formed between two primary sample individuals represent about 5 percent of the formation rates in the population. Column 3 presents the relationship formation rates estimated using the sample of couples identified when non-primary sample individuals are included. For each age range, these rates are estimated as the ratios between the number of primary-sample individuals involved in relationships identified using the algorithm allowing for non-primary sample individuals discussed in Section 3 and the total number of primary sample individuals. Separation rates are estimated as the ratio between the number of separated couples and the number of all relationships that survived through the first, second, and fourth years.

Table A2: Demographic Characteristics at the Time of Relationship Formation

	CCP data (1)	PSID data (2)	Placebo sample of couples (3)
Individual level characteristics			
Average age	36.7	33.5	36.1
Age difference	3.6	3.8	3.7
Age correlation	0.85	0.86	0.82
Census block group level characteristics			
% White correlation	0.63	0.66	0.01
% College degree Correlation	0.48	0.31	-0.00
Median Income Correlation	0.35	0.38	0.02

Note: Column 1 presents the results corresponding to committed relationships among two primary sample individuals. Column 2 presents the results of the couples observed in the PSID data who are between the ages of 20 and 55. Column 3 presents the results of the placebo couples who are randomly matched in the CCP data subject to the same age and age differential restrictions. Age in the CCP data is calculated using the year-of-birth variable. Race, education, and median income variables are the census block group level statistics in the 2000 U.S. Census.

Table A3: Validation Analysis of Using Block-Group Proxies in Regressions

	(1)	(2)	(3)	(4)
Age	0.344***	0.470***	0.421***	0.455***
	(0.027)	(0.029)	(0.028)	(0.028)
Age^2	-0.785***	-1.136***	-1.000***	-1.094***
	(0.079)	(0.084)	(0.080)	(0.082)
Age^3	0.604***	0.919***	0.795***	0.882***
	(0.073)	(0.078)	(0.074)	(0.076)
Individual Char.				
White	0.258***	0.335***	0.216***	0.324***
	(0.008)	(0.008)	(0.008)	(0.008)
Below HS	-0.293***			
	(0.011)			
College	0.384***			
	(0.009)			
Block-group proxies				
Share(Below HS)			-0.499***	
			(0.034)	
Share(College)			1.105***	
			(0.032)	
County proxies				
Share(Below HS)				-0.359***
				(0.055)
Share(College)				1.491***
				(0.051)
Year F.E.	Yes	Yes	Yes	Yes
R^2	0.278	0.175	0.255	0.214
N	22,273	22,273	22,273	22,273

Note: Column 1 presents the estimates of regressing log labor income on observed racial and educational characteristics—a white dummy, a below high school dummy, and a dummy of college education. Column 2 presents the results excluding education attaintment dummies. Column 3 presents the results estimated using census block group-level proxies of educational attainments. Column 4 present the results estimated using county-level proxies. All models control for a third-order age polynomial and year fixed effects. *** denotes the estimate is statistically significant at the 99 percent level.

Table A4: Robustness Tests

	Primary sample individuals only					
	Credit score percentiles		scores 8 Qs before		Living sep. 16 Qs before	
	year 3 or 4	year 5 or 6	year 3 or 4	year 5 or 6	year 3 or 4	year 5 or 6
	(1)	(2)	(3)	(4)	(5)	(6)
$\frac{Initial\ diff}{100}$	1.291***	1.158***	1.163***	1.114***	1.215***	1.105***
	(0.021)	(0.027)	(0.019)	(0.026)	(0.025)	(0.035)
$\frac{Initial\ score}{100}$	0.572***	0.616***	0.699***	0.700***	0.661***	0.653***
	(0.013)	(0.019)	(0.014)	(0.020)	(0.016)	(0.025)
Controlling for						
Age polynomials	yes	yes	yes	yes	yes	yes
Age diff.	yes	yes	yes	yes	yes	yes
Current community char.	yes	yes	yes	yes	yes	yes
Community div. pop. share	yes	yes	yes	yes	yes	yes
Orig. community char. diff.	yes	yes	yes	yes	yes	yes
Yearly FE	yes	yes	yes	yes	yes	yes
State FE	yes	yes	yes	yes	yes	yes
N	29,188	20,518	28,345	19,974	14,516	8,618

	Including non-primary sample consumers			
	Whole sample		Having joint accounts ever	
	year 3 or 4	year 5 or 6	year 3 or 4	year 5 or 6
	(7)	(8)	(9)	(10)
$\frac{Initial\ diff}{100}$	1.278***	1.170***	1.062***	1.026***
	(0.003)	(0.004)	(0.005)	(0.005)
$\frac{Initial\ score}{100}$	0.647***	0.661***	0.687***	0.723***
	(0.002)	(0.003)	(0.003)	(0.004)
Controlling for				
Age polynomials	yes	yes	yes	yes
Age diff.	yes	yes	yes	yes
Current community char.	yes	yes	yes	yes
Community div. pop. share	yes	yes	yes	yes
Orig. community char. diff.	No	No	No	No
Yearly FE	yes	yes	yes	yes
State FE	yes	yes	yes	yes
N	1,195,160	820,103	755,210	588,395

Note. The table reports odds ratios of the logistic regressions associated with a one-standard-deviation change of credit score differentials and average levels at the time of relationship formation. Standard errors are reported in parentheses. *** denotes the estimate is statistically significant at the 99 percent level. The estimations control for a third-order age polynomial of each spouse and the couple age differential; The estimations control for U.S. Census statistics of race, education, income, and share of divorced population of the census block group where the couple lived as of the end of the first, second, and fourth year of their relationship, respectively, for the three sets of results. For results of column 1–8, the estimations also control variables the differentials of these statistics between the census block groups where each spouse lived prior to forming the relationship. Finally, the estimations control for yearly and state fixed effects. See text for details of each of the robustness test specifications.