STOCKHOLM SCHOOL OF ECONOMICS
Department of Economics
659 Degree project in economics
Spring 2020

The Role of Online Dating in Intra-couple Gender Equality

Lovisa Qvarner (24051)

Abstract: Online dating has earned a key role in couple formation over the last two decades. This thesis explores gender differences in sorting across online and offline dating markets with a historical dimension, and thereby includes the aspect of development of market environments over time. I focus on intra-couple gender equality characteristics in couples formed in different matching markets. I propose a framework that explains how couple formation may differ in online and offline markets, building on matching market theory and national demographic balance. The framework suggests that participants of otherwise thin markets have the most to gain from using online dating markets, and highly educated women together with less educated men currently face especially thin offline dating markets in the Unites States. I use two nationally representative datasets on US couples and a difference-in-difference approach to empirically test the hypothesis that an online environment further breeds couples with a different type of gender inequality balance, compared to couples formed offline, where females to a greater extent exhibit higher income and educational attainment relative to their partner. The results partly confirm the expected pattern empirically, where sampled couples that met online are more likely to exhibit a female with relatively higher income than her male partner and less likely to exhibit a male with higher income than his female partner, compared to couples that were formed in an offline matching market. The result provides insight regarding the quality of current market design in online dating markets, demographic challenges and intra-generational mobility.

Keywords: Gender equality, Marital sorting, Online dating, Market design, Man deficit, United States
JEL: D47, D63, J12

Supervisor: Anna Dreber Almenberg
Date submitted: May 13, 2020
Date examined: May 19, 2020
Discussant: John Bolmgren
Examiner: Johanna Wallenius

Table of contents

Table of figures

Acknowledgements

Firstly, I would like to thank my supervisor Anna Dreber Almenberg for providing crucial inputs and feedback regarding the research area, my analysis and my writing as such. Without her guidance, patience and support I would have struggled during this process. Secondly, I would like to thank Felix Schafmeister for sharing his knowledge and providing support regarding the econometrical approach.

Finally, I wish to thank family and friends with whom I have shared my research and gained valuable feedback and inspiration. Especially my boyfriend Richard Valleryd, for his dedication and continuous support in this project.

1 Introduction

The *market* for potential romantic partners is a natural phenomenon that has been present for as long as beings have needed a mate, and most of us have at some point been participants. It is a marketplace called a *matching market*.

Arguably, this matching market is messy, and actions of participants might seem random. Still, by studying its underlying mechanisms and structures through the lens of economics, the economist can understand more about its driving forces. There is undeniably supply of, and demand for, romantic partners and matches between individuals happen each and every second. This type of market is the antipode of a classic commodity market where goods are identical, and an equilibrium price clears the market. Instead, participants are highly engaged in their selection and the market never completely clears. The marketplace becomes a system of agents constantly searching, evaluating and choosing between options. Finally, a participant ends up in a stable match with a partner or decides that it is better to be alone, at least until the participant decides to reenter the market and the algorithm starts over again. This kind of market is part of the recently emerging literature on matching markets and market design. Studies of two-sided matching and its inherent game theory was largely initiated by Gale and Shapley (1962) and more recently further studied by, for example, Nobel Laureate economist Roth (2002; 2004; 2008; 2015) through the lens of market design.

As human societies and lifestyles evolved during the course of history, the partnership markets' structures stayed relatively similar and experienced only relatively minor design changes before the emergence of the internet.[1] Agents and stakeholders interacted with each other in their immediate communities, evaluated each other, and eventually, agents chose or were assigned with a partner based on availability and a set of desirable attributes. It is only in the last few decades that technology has enabled a scalable and potentially efficient environment for matchmaking, and the market participants are undeniably using it. Almost 40 percent of US adults met their partner online as of 2017, which represent a two-fold increase in less than a decade (Rosenfeld et al. 2019). Today, participants can enter the market without leaving the comfort of their homes, scroll through a thousand profiles of suitable partners and even initiate contact. The market would seem to have been designed towards efficiency. However, what kind of couples arise in this

[1] Pre-internet dating market design efforts include human match makers (usually religious leaders or elderly women of the society), personal advertisements since the rise of modern newspapers (early 1700s) and later, video-dating (1980s). However, the use of the latter two approaches never became mainstream. (Finkel et al. 2012)

environment? I choose to study a relatively narrow aspect of couple characteristics—intra-couple gender equality, which will be studied in two dimensions. The purpose is to investigate intra-couple gender equality in sorting patterns generated by online and offline markets, respectively. To investigate potential differences in couples formed online versus offline, I look into generated sorting patterns by each market. More specifically, I answer the question whether couples that were formed online exhibit a different internal gender equality balance compared to couples that were formed offline, in terms of educational attainment, maternal educational attainment and income distribution. In addition, the purpose is also to study the intra-couple gender equality balance trend across markets *over time*, as potential change has taken place regarding the design of online market structures, the self-selection into different market types and the national socio-demographic. The geographical focus lies on the United States. In the background section, I marry the market design literature with current gender differences in US socio-demographics to arrive at a set of hypotheses.

The economic implications of these questions are important and has caught the interest of scholars and the general public for decades. Following the work of Becker (1973;1974), numerous studies has aimed to provide clarity on how populations sort in mating, including the potential implications for inter- and intragenerational mobility and inequality in society. The answers have, and has always had, a key role to understanding the socioeconomic characteristics of the society our descendants will live in.

I begin this thesis by presenting relatively recent developments in the market design area, together with more practical topics regarding the online market environment as such and the socio-demographic balance of the United States. Then, I move on to present previous research on assortative mating, preferences and behavior in online dating settings. I then test my hypotheses using two datasets from Stanford University that have primarily been investigated by sociologists (two waves of the *How couples meet and stay together,* HCMST, survey). The first sample was collected in 2009 (n=4,002) and contains data about how adult Americans met their spouses or romantic partners and characteristics of the couples. The same respondents have since been interviewed four more times to follow-up on the relationship's progression during the following six years. In 2017, a second fresh sample (n=3,510) was collected and similar questions were asked. I use the 2009 and 2017 samples exclusively in this thesis, as they contain no overlap in respondents and are both considered nationally representative. As I study gender equality differences, I look at opposite-sex couples exclusively, and thus exclude all same-sex couples. I also exclude couples that met more than eight years prior to each survey in order to be able to compare more recent sorting behaviors in the two samples and prevent overlap in year which the couple met. The final sample size is 753 observations in the 2009 sample and 704 in the 2017 sample. Several gender equality variables regarding income distribution and education are tested using a modified *difference-in-differences* regression approach, with *time period in which the couple met* and *market in which the couple met* (which can be either offline or online) as explaining variables in the regression analysis. I execute Wald tests to distinguish additional patterns, that are not directly displayed in the regression output. The estimation model is not meant to establish any causal relationships, only to compare couple characteristics across market and time period.

I compare couple characteristics by observing differences between the male and the female of each couple, more specifically: income distribution, educational difference and educational difference of both partners' mothers. I find statistically significant intra-couple equality differences between couples initiated online and offline in the 2017 sample, and an interesting development for the gender equality variables over time. Among couples in the 2017 sample, women are more likely to have higher income than their male partners if the couple met online compared to couples that met offline, although, I find no statistically significant result that they are more or less likely to exhibit income equality. In the 2009 sample, I find no statistically significant income differences across market. For the inequality of having a female with higher income, a differential effect for meeting online in the 2017 sample compared to the 2009 sample is found. These results indicate that sorting from the two market types are different in terms of intra-couple gender equality in 2017, and that the differences has increased over time, with couples displaying a higher likelihood of having a higher income female if they met online during 2009-2017 as compared to 2001-2009. Moreover, I find no statistically significant differences in any educational difference across market type nor time period.

As inherent with the study design, no causal relationships between means of meeting and couple characteristics can be drawn due to biased self-selection between market types. Although, the result of the study is useful in explaining the role of online dating in society and is previously unrepresented in the research landscape.

This thesis continues as follows: in the next section, I discuss the literature on matching markets, market design and the dating markets' environment. I thereafter present previous research regarding behavior, preferences and sorting in online dating. I subsequently summarize the material in a model framework, from which my hypotheses are drawn. Then, I discuss the contribution of my study and present the research questions, method, dataset. Finally, I present the results along with a discussion of robustness, implications, main issues and possible extensions.

2 Background

This section begins with an introduction regarding the market type in which matchmaking occur– the *matching market*, together with its three critical design aspects as proposed by Roth (Roth 2008, p.286). Thereafter, market structures and considerations regarding the dating markets specifically are presented. In this section, the emergence of online dating market structures is summarized, and the online dating markets are evaluated using the three critical design aspects. Lastly, I summarize the national socio-demographic gender balance that all dating markets operate in.

2.1 Matching markets

2.1.1 Characteristics

Matching markets are a special kind of marketplace where participants must be *matched* with each other in order to trade, and the 'goods' traded are characterized as private, indivisible and heterogenous. Roth describes them accordingly:

> "Matching is economist speak for *how we get the many things we choose in life that must also choose us* [emphasis added]." (Roth 2015, p.16)

For example, a boy asking a girl for date—the boy chooses what girl to pursue, although to be successful, the girl must choose him back. The *good*, a relationship, is private (rivalrous and excludable in nature), highly differentiated (since a relationship with one girl would be different from a relationship with another) and indivisible (cannot be shared with others). Another example of a matching market is the labor market.

Sometimes there are structures available to support the matching market, such as an application and selection process or a dating website. The structure itself and the way agents handle the process determine the entire matching outcome (Roth 2015). These ever-present structures are examples of market design, a way in which the designer assist the participants of the market to match appropriately and efficiently.

A distinctively different characteristic of matching markets compared to other markets in society is the price mechanism. Anyone that has enough money can buy a commodity from a commodity market, there is no need for applications or sales pitches—the price mechanism brings individuals together at the price where supply equals demand. On the other hand, while prices may be present in some matching markets, the price mechanism works in a different way. Consider an attractive employer: she does not lower wages until there are just enough interested unemployed individuals left in the market, she rather prefers the most qualitied workers and pay them accordingly (Roth 2015, p.17). As it is not the price that makes the participants come together in a matching market, a market structure that enables efficient matchmaking is needed instead.

2.1.2 Underlying mechanisms

Gale and Shapley (1962) evaluated the mechanisms of this market type relatively early. They were the first to propose a stable market outcome as when agents are matched in such a way that switching partners (assuming that both partners of the new couple agrees) or choosing no partner at all would not make any participant better off. Additionally, they proposed the *deferred acceptance algorithm*, a simplified system of algorithmic behaviors in the marketplace.[2] They proved that the algorithm always produce stable matches in simplified settings and their findings still largely define the market design field today.

The deferred acceptance algorithm is modeled accordingly:

1) Individual agents on both sides of the market (men and women) has set preferences in the form of a ranking over the agents of the opposing side.
2) Agents from one side (e.g. the men) extend offers to their most preferred agent on the opposite side of the market (e.g. the women). Each receiver can get several, one, or no offers.
3) The offers are reviewed by the receivers (the women), and each of the women holds on to their highest-ranking offer while the others are rejected.
4) The agents whose extended offers were rejected (rejected men) move on to propose to the next candidate (women) on their list, even if that candidate already holds on to an offer. Then, the receivers (women) review all present offers again, and once again holds on to the best one.

[2] In fact, the same algorithm was used ten years earlier in a medical labor market clearing house, on the initiative of Harvey Hendren whom noticed a flaw in the design of existing clearing houses. Although, the algorithm was not properly documented (Roth 2015, p. 150).

The final step is repeated until no further offers are made, the deferred offers are now accepted, the market is cleared, and stable matches are produced. Agents can end up alone—as they will stop extending offers (or stop tentatively accept offers) if they consider it better to be alone than to be matched with any of the individuals left in the market. In addition, the algorithm can provide several sets of stable matches and thereby indicates a *two-sided matching problem*, where the agents on the proposing side of the market are in control and create the most favorable matching outcome for themselves.

As of today, a fully cleared and stable dating market in the Gale-Shapely fashion is not achieved in any market setting. The 'offers' are extended under more asymmetrical information, in an unstructured environment and from both sides simultaneously. The cost of search and evaluation are high, and agents are forced to repeat the matching process to uncover hidden information with different mates. Still, much of the research conducted in this field is based on the foundation laid by Gale and Shapely.[3]

2.1.3 Three critical design aspects

Arguably, with a perfect market structure, the dating market might accommodate participants to act according with a proposed algorithm and enable the market to become better in terms of efficiency and stability of matches. Indeed, the market structure can be modified, and Roth describes it accordingly:

> "The economic environment evolves, but it is also *designed.* [emphasis added]" (Roth 2002, p.1341).

The various matchings take place in different market settings (e.g. LinkedIn for the skilled worker and a bar or dating site for the singles). These are the kinds of marketplaces that Roth and colleagues design to work better in terms of efficiency and stability. For example, by collecting data on kidney donor-patient pairs that were not a medical match, they built a marketplace where medically correct matches could be made between individuals on the patient and donor sides, respectively (Roth and Sonmez 2004). The solution was to create a donor chain. Before the implementation of the new design, the matching market was failing and there was not an efficient way for participants or facilitators to find and evaluate potential matches.

[3] See for example Hitsch et al. (2010) for a study specifically regarding online dating which is based on the deferred acceptance algorithm.

Matching markets can potentially be designed to further efficiency and Roth identifies three kinds of market failure that market design aim to eliminate (Roth 2008, p.286).

1) Markets are too *thin*: there are not enough participants to achieve appropriate matching. Markets need to achieve proper *thickness* to function efficiently.
2) Markets become *congested* with thickness: participants cannot evaluate opportunities efficiently as a result of a crowded market with subpar structures. Markets need to have a smart structure that allows participants to evaluate options in reasonable time.
3) Failure to provide *safety* for participants to reveal or act on information. Markets need to have a safe environment that enables efficient information exchange and thereby reduces informational asymmetries.

The three aspects of market properties can be used to assess whether a matching market is well-functioning or not. The next section connects the three design principles to market design in dating, but first provides an historical overview of the structural development in online dating.

2.2 Dating markets

2.2.1 History

Commercial and social institutions have been facilitating courtship and marriage long before the emergence of the internet, with for example human matchmakers, newspaper advertisements, and video-dating, additionally, computer power has been used in matchmaking for over 70 years (Finkel et al. 2012). Since then, the online dating market has emerged, and this section will focus on the last two decades of development.

Together with the popularization of the internet and the development of personal computers, new types of dating services were founded. The modern approaches of online dating services can be sorted into three generations[4]: 1) online personal advertisement sites; 2) algorithm based matching sites and; 3) smartphone-based dating applications. Finkel et al. (2012) concludes that the first major actor in the online landscape was Match (formerly match.com), that launched in 1995 and quickly gained popularity. Several imitators followed match.com and launched their own sites in the following years, among which many were niche sites that served specific subgroups.[5] They describe that these sites essentially functioned as search engines where the users could browse

[4] Following the categorization proposed by Finkel et al. (2012).
[5] For example, specific sites for age groups (e.g., SeniorPeopleMeet), religious orientation (e.g., JDate) and social status (e.g., DateHarvard).

among online personal profiles and post their own profile, and that the revenue structures usually were based on membership fees or on-site advertisements.

The second generation was initiated by eHarmony in 2000 and was promoted as ‘science-based’, several similar competitors entered the market over the following years. Most often, singles would provide the service with information and be matched with potential partners according to the sites’ compatibility algorithm, and services often hired social or behavioral scientists to support the matching process (Finkel et al. 2012). The second generation often charged users a higher membership fee compared to the first generation (Finkel et al. 2012). There were some further differentiated sites of this generation that claimed to use genetic and immunological compatibility as the foundation of their matching process, with GenePartner (launched 2008) and ScientificMatch (launched 2007) being two examples.

With time, the approaches of self-selection and algorithm-selection converged, and sites of the first generation soon included algorithm suggestions to their otherwise self-browsing site (Gelles 2011). The third and latest generation emerged in 2008, as Apple launched their app-store where independent companies could provide applications for the smart phone. The app-based dating services are often location-based and typically consist of personal profiles with limited information and a few photos (Finkel et al. 2012). Users can often link other social media accounts to their profile to provide more information. These apps are often ‘gamified’ where users categorize potential partners into “like” and “don’t like” based on geographical proximity, looks and the limited information provided. What potential partners that are displayed to a user is typically based on the user’s activity, geographical proximity, relative age and gender but also on the characteristics of profiles that the user likes and the characteristics of profiles that like the user (Tinder 2019).

Since the availability of personal computers increased in the late 1990s until today, online dating has grown from almost non-existent to the most common way of meeting. Meeting online surpassed the previous dominant way of meeting (through a friend) around 2013 (Rosenfeld et al. 2019).

2.2.2 Market design evaluation

This section evaluates the properties and mechanisms of online markets compared to offline markets in the light of the three critical design aspects proposed by Roth (Roth 2008 p. 286).

Thickness

The condition of *thickness* is improved by decentralized online market structures. The sets of potential partners connected to various dating services and applications are larger than the sets of potential partners connected to, for example, one's mother or one's friend (Rosenfeld et al. 2019).

Large choice sets are valuable to everyone engaged in search (Rosenfield 2017), and especially valuable to participants looking for more unusual characteristics in a partner (Rosenfeld and Thomas 2012). Individuals of otherwise thin markets find relatively higher value in online dating as their market thickens significantly. Agents can access a decentralized online marketplace and are no longer limited by geography, their social connections and otherwise hidden information, such as for example, sexuality.

Congestion

The success in *Thickness* can become a downfall in *Congestion*. Congestion regards the inability of participants to efficiently evaluate options in a reasonable timeframe, and the online market can potentially get congested. I find two major theories behind why the online market experience congestion.

Firstly, the online market being too 'safe' can cause unnatural behavior in approaching potential partners. In social psychology, a proposed mechanism called *the matching hypothesis* (Walster et al. 1966) claims that participants are strategic in their mate selection, namely, individuals pursue prospects that mirror their own social desirability since these efforts are more likely to end in success. Although, regarding interactions in online dating, studies suggest the opposite behavior: participants are likely to approach the most socially desirable counterparts regardless of their own desirability (Hitsch et al. 2010; Kreager et al. 2014).[6] As a result, socially desirable individuals receive plenty of requests while less socially desirable individuals stand empty handed. Kreager et al. (2014) propose the explanation that the online environment drastically reduces negative consequences of rejection and offers a too 'safe' way to initiate contact. They argue that an explanation to why the matching hypothesis often holds empirically for marriages but is not visible in online dating behavior is that most online interactions do not transform into actual relationships or marriages (Kreager et al. 2014).

[6] These findings resemble behaviors suggested by the Gale and Shapely (1962), in the Gale-Shapely deferred acceptance algorithm.

Secondly, the imbalanced flow of information could have an additional/complimentary explanation in form of informational asymmetries regarding the distribution of participants. There is evidence that participants change their behavior as market thickness varies, specifically, more available choices affect participants to become picky and reluctant to settle. Fong (2019) finds, both empirically and experimentally, that studied participants in online dating markets express increasing selectivity as they experience an increase of potential partners. Vice versa, they become less selective when they observe hard competition for the potential partners available. When accounting for this *thickness responsive selectivity*, Fong found that matching exhibits decreasing returns to scale. In other words, the bigger the market—the relatively fewer matches. Fong's findings indicate that participants in dating sites might behave unfavorably picky when they cannot observe the distribution of other participants of the same side. While daters in bar, for example, can transparently observe the competition and act accordingly. This is a fact that gives rise to a major problem, as online dating services often advertise the abundance of potential partners in the marketplace and do not provide indications of the extent of competition from the participant's own side.

As a response to experienced inefficiencies, some online dating services have managed to reduce this type of congestion by introducing a preference signaling mechanism. By enabling participants to send virtual tokens that are made scarce in nature, the receiver can observe the sender's genuine interest and is able to better sort through the overflow of shallow information, as a result, the match success rate increase for the sender and the market produces more matches in total (Lee and Niederle 2015).[7]

To conclude, the current online market structure creates a reduced sense of rejection and seldom provides ability to observe competition, thereby influencing participants to behave unnaturally in their search and evaluation phase. These characteristics of the market structure combined with a high number of participants generate a massive flow of information directed towards popular agents, creating imbalanced flows of information compared to offline markets. The imbalanced informational flows incapacitate the popular participants from evaluating efficiently and the less popular participants are not given a fair chance. There is also a frustrating set-up for failure *after* the online match has been made, as participants try to accurately evaluate the potential dates that

[7] Arguably, a feature that could be seen as to rather alleviate the symptoms than handling the causes of the congestion. Now, less socially desirable agents can choose to send tokens to signal interest and desirable agents can observe that. A procedure that probably will not reduce the full congestion caused by unnaturally brave approaching and thickness dependent selectivity.

they met in a setting of informational asymmetry (Meltzer 2016). Indeed, there seems to be a problem with hidden and skewed information in the online marketplaces, a fact that further incapacitates participants to efficiently evaluate options.

Safety

Users of online dating services are frustrated about other users providing misleading information, while they also experience their own privacy concerns (Meltzer 2016). This indicates that there are flaws present that prevent participants from safely revealing and acting on information, and thereby are unable to enjoy efficient informational exchanges. A 2019 survey by Pew Research Center (2020) concluded that over 70 percent of online daters think that it is very common for fellow users to lie to appear more desirable, and 53 percent of women felt that online dating is an unsafe way to meet people. The perception of online daters has been confirmed in several studies—deception is indeed frequently observed in online dating settings (Markowitz and Hancock 2013; Drouin et al. 2016). For example, by comparing users' online profiles to their confirmed personal data, Hancock et al. (2007) found that up to almost 60 percent of sampled US adults lied about their appearance (regarding height, weight and age).

Additionally, online dating platforms are used by criminals as a way to find victims. The Federal Bureau of Investigation (2020) reports that over 4.75 billion dollars were reported stolen online in *Romance and confidence fraud* in the US during 2019 and this trend is reportedly increasing exponentially. Evidently, online dating services often fail to provide a safe and efficient environment for participants to trust each other and to provide and act on information. The information exchanged is instead often skewed, and there are direct threats present in the form of criminals looking for targets. Serious informational asymmetries are present in online marketplaces.

Finally, evaluating the achievements of the online dating markets in the light of Roth's three conditions for a smooth matching market, it seems as the market achieves great thickness, although has not yet solved arising problems regarding congestion and safety. In fact, there are several identified flaws in the market design that hinders efficient evaluation, mainly due to informational asymmetries and safety concerns. The identified success and flaws are shared between the main different types of online dating markets presented in section *2.2.1 History*, as they share the features discussed in this section.

2.3 The 'man deficit'

The final aspect of the background section deals with the actual participants present in the marketplaces. The markets can only produce matches containing the characteristics of the in-market singles, making this section the final piece of the model framework. The 'man deficit' is a term that has been gaining popularity over the last years, with several newspaper articles discussing the phenomenon.[8] It refers to a pattern of gendered socio-demographic differences in the US which has emerged since females surpassed males in college completion in 1980s.[9]

Looking at assortative mating in the United States, regardless of matching market, researchers most often find positive assortative mating, where men and women of similar socio-economic class marry (Eika et al. 2014; Greenwood et al. 2014). However, there are indications that the market participant balance of US dating markets is unproportionally unfavorable for highly educated and aging women, there are too many single women of high socio-economic class relative to the amount of socially desirable single men available. Indicating that these women face a thin dating market with high competition. A survey of US adults by Wang and Parker (2014) found that almost 80 percent of surveyed women claimed that it is very important to find a spouse with a steady job (only 46 percent of men agreed), while for every 100 unmarried women there were only 65 employed unmarried men. A similar pattern was found regarding education, for men, the higher the degree—the more likely they are to be or have been married, only 14 percent of sampled men with a post-graduate degree had never married, while the number for men with a high school degree or less was 25 percent. Women express a reverse relationship, the higher degree—the less likely to be or to have been married, only 16 percent of women with a high-school degree or less had never married while 18 and 20 percent of women with post-graduate and bachelor's degrees respectively, had never married. At the same time, women continue to outpace men in college completion in the United states, women earned over 57 percent of all awarded bachelor's degrees and 62 percent of all awarded master's degrees in 2017 (NCES 2019). This gendered educational gap diverged from zero in the 1980s and has been increasing since. Wand and Parker (2014) found that age is another important demographic factor: at age 25, unmarried men is surplus to unmarried women, with 118 men to every 100 women. The ratio declines with age and at age 64, there are only 62 unmarried men to every 100 unmarried women. Passing the 1-to-1 balance at age 40. The explanation behind the numbers is that women (especially over the age of 45) are much more likely

[8] See for example the article 'Broke men are hurting American women's marriage prospects' in the New York Post (Frishberg 2019).

[9] Females surpassed males in number of earned bachelor's degrees in 1981 and master's degrees in 1986, the gap in college completion has been increasing since (NCES 2019).

to be currently divorced or separated, while the older divorced men more quickly/often find a new partner. Looking at adults over the age of 45, 37 percent of women and 22 percent of men are currently are divorced, separated or widowed as they find younger women.

Looking at the US today, the imbalanced demographics have been be translated into direct implications for the marriage market. By studying couple characteristics of newlyweds, Lichter et al. (2020) estimated the sociodemographic characteristics of unmarried women's potential spouses. They compared the husbands of women with certain social desirability and predicted the level of social desirability of potential husbands to the unwed women, then they compared the potential husbands to the men available in market. The potential husband's income would be almost 60% higher than that of the available men in the market, he would also be 30% more likely to be employed and almost 20% more likely to have a college degree. The researchers conclude that the deficit of 'suitable' men implicate that US women might choose to stay single or marry less suitable men. The same pattern is visible in other modern countries such as Japan (Raymo and Iwasawa 2005).

Jon Birger, the author of 'Date-onomics: How Dating Became a Lopsided Numbers Game', is an advocate of the theory that the man deficit is fundamentally changing behavior in the dating markets (Birger 2015). The scarcity of highly educated men would give—the now very attractive and few men— an incentive to delay marriage and instead create a hook-up dating culture. Birger predicts that professional women and men of lower education will have to start finding each other in the dating market, simply because highly educated men are scarce and reluctant to settle, and women of low education are few. Birger was not the first to propose that these kind of macro-level ratios could affect micro-level behavior[10], although, the simplicity of these theories has been questioned.

3 Previous research

Previous findings regarding intra-couple gender equality in sorting patterns across dating markets are sparse. There are several studies that in a few ways contribute to the area, although most studies investigate behavior and preferences in online dating, not lasting sorting outcomes (like relationships or marriages). The few studies that capture the link from online interactions to sorting in society focus mainly on assortative mating, a topic that is only partly related to my research

[10] See the work of psychologist Marcia Guttentag (1983).

question. The assortative mating literature often focuses on individuals of a couple without considering the partners' respective gender. *Endogamy*, when individuals marry within a social group, and *exogamy*, when individuals marry outside their social group are the main aspects studied in this field. The scope does not necessarily include intra-couple gender differences.

Firstly, I present the findings of a typical approach chosen regarding behavior, preferences and matching in online dating, specifically, using data collected from dating sites and modeling matching outcomes. These studies focus on the early phase of couple formation, within specific dating services. Online daters from Boston and San Diego were studied by a research team in 2010 (Hitsch et al. 2010). The researchers used data on user attributes and interactions collected directly from an online dating site to predict sorting outcomes, based on the Gale-Shapley deferred acceptance algorithm. By analyzing the interactions of agents combined with the attributes of whom they interact with, general mate preferences were estimated. Thereafter, the deferred acceptance algorithm was applied to predict stable matches. The predicted matches resembled the actual matches produced on the dating site, and the actual matches were approximately stable in a Gale-Shapely fashion. The researchers interpreted this result as an indication that search frictions were low and that the design of the site was efficient. The positive assortative mating patterns visible in the sample was considered to arise as a result of preferences and the market mechanism, not search frictions. Additionally, the researchers reweighted the sample to resemble the full population and performed the analysis again. This time they compared the predicted stable matches to actual marital patterns in society to assess whether the same sorting pattern, which was considered preference driven, could be found in actual marriages—which it was only to some extent. The results indicated that the sorting of the online market was different from the sorting in marriages and the researchers hypothesize higher search frictions offline to be the reason why.[11] Regarding mate preferences, they found that women care about their potential partners' income level about twice as much as men, in the sense that they want the men to earn a substantial amount.[12] Regarding education, both genders preferred a partner of similar educational level. Although, while women had a preference for men of similar or higher education, men had a tendency to avoid women of relatively higher education. In a typical match of the site, men would earn $49,000 more in annual income and have spent half a year more in school.[13] A meta-review

[11] Although, it could also be due to different search and evaluation behavior online, as studied by Fong (2019) and Kreager et al. (2014) and presented under *Congestion*.
[12] As is consistent with the findings of Wang and Parker (2014), presented under 2.3 The man deficit.
[13] The matches only represent online matchings and are thus not necessarily connected to a relationships or marriages, individuals can be included in several matches.

of studies regarding gendered differences in online dating behavior conducted by Abramova et al. (2016), provides support for the findings of Hitsch et al.. It was concluded that several studies confirmed the gendered pattern in educational preferences and also that women, in contrast to the men, considered a potential partner's income a very important aspect. Several studies suggested that women most often preferred a partner of high socio-economic status, while the results regarding men were ambiguous.

I move on to present the sparse findings connecting *specific* online interactions to societal sorting. Lee (2016) manages to execute a similar approach as above but also to connect the findings to sorting in offline couples, by observing the couples again after they migrate offline. Lee used a novel dataset with verified information on South Korean couples that met on an online dating site and then got married. Lee found that the online daters were more likely to find someone of similar occupation or industry (further endogamy) but less likely to marry someone with similar educational level (further exogamy) compared to couples that met offline. It was concluded that the observed patterns suggested that different marital sorting patterns arose from the online dating service in question compared to couples that met offline.

Finally, I present the findings from a fully aggregated approach of instead using a large random sample of the population and study sorting outcomes combined with reported means of meeting, even if the results are only partly related to intra-couple gender equality. Thomas, one of the researches behind the HCMST datasets has conducted a study regarding exogamy in online dating for US couples (Thomas 2020). Thomas used data from two waves of the HCMST survey (2009 and 2017) and excluded couples that were formed before the emergence of the online market (1995 was arbitrarily chosen), then he differentiated the couples by market in which they met and conducted the analysis while controlling for the diversity of the respondents geographical location. Regarding education, Thomas found that couples that met through online dating express significantly higher probability of educational exogamy compared to couples that met offline.[14] Additionally, a similar study (Potarca 2017) using both US and German data found mixed results on the effect of meeting online on assortative mating. Only differences between meeting online and specific offline contexts that usually fosters endogamy (such as school, family, friends and religious venues) was found. The online environment was associated with weaker endogamy compared to the specific offline contexts.

[14] Thomas's (2020) findings are consistent with Lee's (2016) findings regarding educational exogamy in South Korea.

4 Outlook and contribution

The background and previous research sections combined provide a foundation for this study, that can be used to model how online markets potentially differ from offline markets, and what sorting patterns should be expected as the result. The constructed model can, in turn, provide a theory on why intra-couple gender equality should differ across market types. Figure 1 represents an illustration of market properties (right hand side) in relation to stage of search and selection process (left hand side).

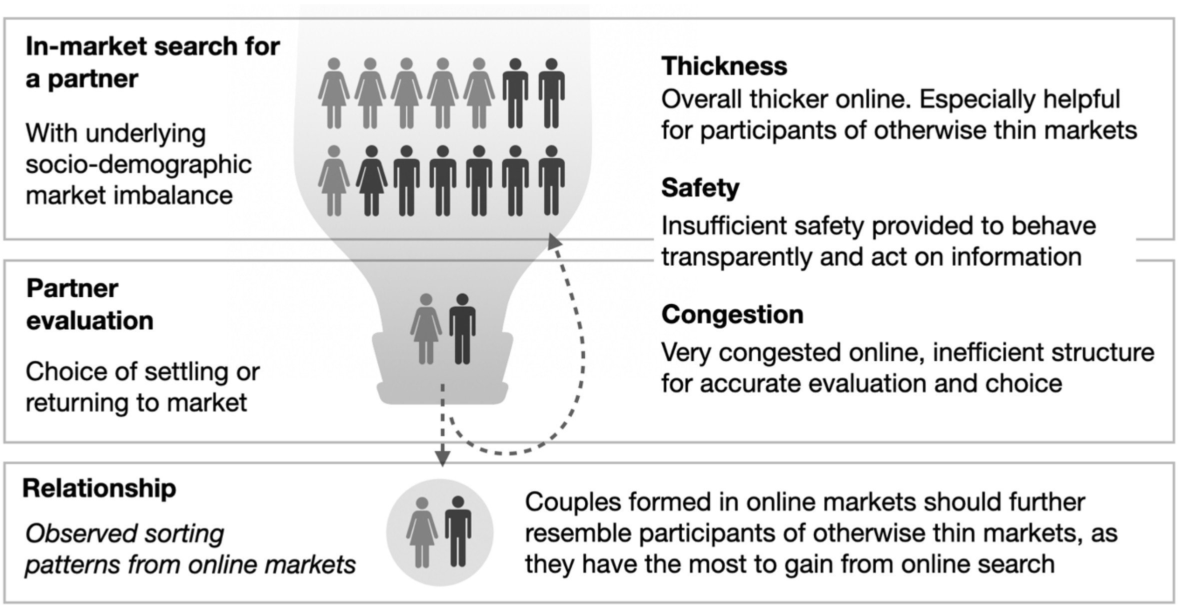

Figure 1. Framework for differences in properties of the online market
Note. Illustration created by the author.

Evaluating the market achievements of the online markets, only the condition of thickness is considered superior compared to offline markets. Considering the full US population, the man deficit is present and affecting the market thickness for subgroups of the population. Namely, that highly educated females and less educated men are cohorts of abundance, while highly educated men and less educated women are scarce. This indicates that highly educated women face hard competition for the educated men and that the men face hard competition over the women of lower education, assuming that women do not want to marry men of a lower social class and men does not want to marry women of a higher social class.[15] As agents of the dating market follow these preferences, women of high education and men of lower education might end up alone or start to disregard their preferences to be able to find a partner.

[15] An implication that several studies in the literature review confirm, it was found that men shy away from women of higher education than themselves and women are drawn to high-income and educated males. (Hitsch et al. 2010; Abramova et al. 2016).

Previous research found that societal sorting patterns from the online market more often exhibits educational exogamy (Lee 2016; Potarca 2017; Thomas 2020), while several studies suggested that both genders prefer educational endogamy or a male of higher education (Abramova et al. 2016). It would be logical to assume that, if the online market has a better design than the offline market, the collectively preferred pattern would be found to a greater extent in couples that were formed online. However, I find serious congestional and safety problems in the online market structure, that indicate a malfunctioning market. Arguably, the main positive contribution of the online market is to increase thickness, which is especially valuable to participants of otherwise thin markets (Rosenfeld and Thomas 2012). I propose that the reason for the findings of further exogamy online is that thin market participants of both sides find each other online, not that the online environment fosters couples closer to the collective preferences. Namely, both men and women can better find partners outside of their social class online, as their social network will not limit the amount and type of potential partners available.

Regardless of the growing interest from both scholars and the general public, we do not yet have a clear understanding about the role of online dating in sorting of marriages and relationships. My view is that the main flaw of previous research consists of an overreliance on the specific dating sites, where data is often collected directly from a dating service and that in itself enables two biases to arise. Firstly, participants are inclined to provide false personal information in online settings (Hancock, Toma et al. 2007; Markowitz and Hancock 2013; Drouin et al. 2016; Pew Research Center 2020), leading to collection of biased data. Secondly, the observed online matching outcomes of the online sites are hard to connect to actual sorting patterns in society, arising as participants move offline and settle down.[16] While these studies have proven that online matching has certain characteristics in terms of participant behavior, they are insufficient in explaining societal sorting patterns.

The studies that overcame these hinders are few and only partly investigate potential intra-couple gender differences across markets. Thomas (2020), Potarca (2017) and Lee (2016) all studied assortative mating in online versus offline market settings and were overall able to find greater negative assortment/exogamy in online couples. Although, the direction of the difference between the male and female as such was left to the imagination. This study aims to provide clarity in this

[16] With Lee's study (2016) as a unique example that overcomes both biases by also collecting data after the couples marry, and only using verified data from the dating site.

gap of the research and explore the direction of the difference from a gender equality point of view.

Like Thomas (2020) and Potarca (2017), I choose the aggregated approach compared to Hitsch et al. (2010) and Lee (2016) for this paper, by leveraging Stanford University's nationally representative samples of US adult couples (HCMST 2009 and 2017) that was used in Thomas's research regarding exogamy (2020). The couples' internal characteristics and their way of meeting are studied to detect potential intra-couple gender equality differences across markets. This way, I manage to completely reduce the dependence on specific dating services (compared to e.g. Hitsch et al. (2010)) and am further able to be certain that my patterns are representative for sorting in the nation (compared to Lee (2016)). Additionally, I study a topic that has previously been unexplored in an online market setting.

5 Research question

The research questions of this thesis are as follows:

1) Do couples that were formed online exhibit a *different* internal gender equality balance compared to couples initiated offline, in terms of educational background and income differences?
2) Has the intra-couple gender equality balance for couples that meet online changed differently over time compared to couples that meet offline?

Based on the model framework for differences in the online market, with matching market theory and socio-demographic balance as a foundation, combined with previous research in the area, hypotheses regarding both questions can be formed.

1) Couples formed online should further exhibit *financially and educationally strong women* relative to their male partner compared to couples formed offline, since these characteristics represent participants of otherwise *thin* markets. Thus, online couples should *not express further equality* compared to couples that formed offline.
2) The couples that were formed online should further exhibit financially and educationally strong women relative to their male partner *over time,* compared to couples that were formed offline. The intra-couple equality balance should somewhat have mirrored the growth of the man deficit, and not be interrupted by better designed structures of the online markets, since the online markets are still generally considered very congested and unsafe.

6 Method

In this section, the study design and its econometric specifications are introduced together with considerations regarding the method.

6.1 General sample and variables

To shed light on the purpose of the thesis, I consider internal educational differences and income distribution for couples formed in the two market types. Figure 2 illustrates a breakdown of the topic into relevant variables. The difference in partners' years of education together with the difference in partners' mothers' years of education represent the relative educational background. Maternal education (one's mother's education) is strongly associated with the child's cognitive development, but also a key predictor of other characteristics within the family that are closely related to children's well-being, such as economic security and family structure (Jackson et al. 2017). In this sense, the education of an individual's mother can resemble part of the social class which the individual is born into. The analysis becomes more nuanced when a predictor of social and economic background is incorporated. Relative income differences are represented by three dummy variables that indicate whether the male earns more, or if the partners earn about the same amount or if the female earns more. The five variables studied (bottom row of figure 2), are interrelated as they all represent different aspects of the social class, economic status and educational background of the individual.

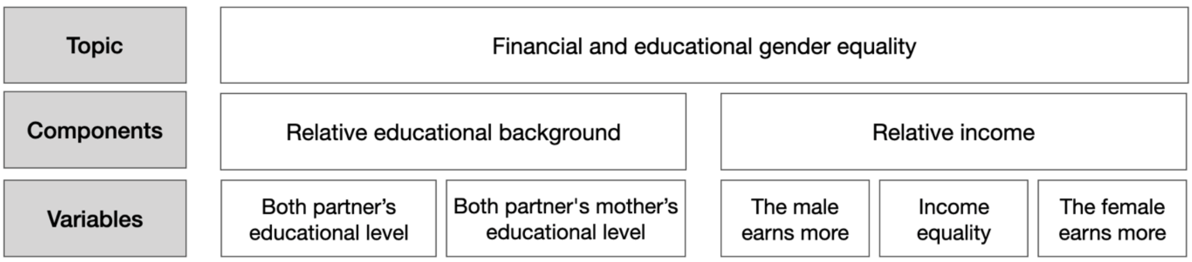

Figure 2. Breakdown of topic into variables studied
Note. Illustration created by the author.

To add an historical dimension to the study, couples formed in two different time periods are studied. The couples are grouped in a systematic fashion according to figure 3. The timing of HCMST surveys support this grouping criteria, as the two fresh samples were collected in 2009 and in 2017. Group 1 and 2 belong to the 2009 sample while group 3 and 4 belong to the 2017 sample. Looking forward, into analysis and results, group 2 and 4 is referred to as *online couples* and group 1 and 3 is referred to as *offline couples*.

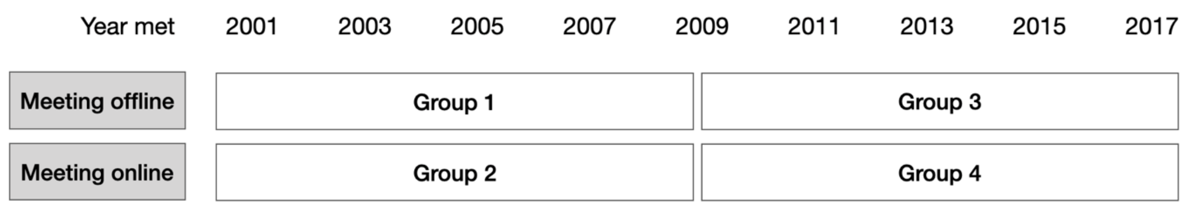

Figure 3. Specification of groups studied
Note. Illustration created by the author.

6.2 Regression model

A *difference-in-differences* variant is used, with which I am able to distinguish patterns across time period and market type. It is not a true *difference-in-differences* model in the sense of establishing a causal relationship. Instead, it is used to compare overall online versus offline couples, to compare the two full samples against each other and to compare whether the trends are different over time for the two markets: whether meeting online has a differential effect in the 2017 sample compared to the 2009 sample. The variables to be tested are put as the dependent variable and the tests are conducted on *couple level*, where each observation represent one relationship or marriage.

$$equalincome_i = \beta_0 + \beta_1\, online_i + \beta_2\, time_i + \beta_3(online_i * time_i) + X_i + \varepsilon_i$$

$$maleincome_i = \beta_0 + \beta_1\, online_i + \beta_2\, time_i + \beta_3(online_i * time_i) + X_i + \varepsilon_i$$

$$femaleincome_i = \beta_0 + \beta_1\, online_i + \beta_2\, time_i + \beta_3(online_i * time_i) + X_i + \varepsilon_i$$

$$educationdiff_i = \beta_0 + \beta_1\, online_i + \beta_2\, time_i + \beta_3(online_i * time_i) + X_i + \varepsilon_i$$

$$motherseducationdiff_i = \beta_0 + \beta_1\, online_i + \beta_2\, time_i + \beta_3(online_i * time_i) + X_i + \varepsilon_i$$

i represent observations; *online* is a dummy that specifies means of meeting, that takes the value 1 if the couple met online and 0 if they met offline; *time* a dummy that specifies what sample the couple belong in, which can be either 2009 or 2017 (by design, the dummy also specifies the time period in which the couples met).

On the left-hand side, *equalincome* is a dummy that indicates whether the partners earns approximately the same amount*; maleincome* is a dummy that indicate whether the couple has a male that earns more than the female and *femaleincome* is a dummy that indicate whether the couple has a female that earns more than the male. The income distributional dummies collectively are aimed to be mutually exclusive and collectively exhaustive for the full sample, meaning that every observation should be attributed only one of the three dummies, and no observation should be left with without a dummy. *educationdiff* represent the gender difference in years of education and *motherseducationdiff* is the gender difference in years of education of the partners' mothers. The educational differences are obtained by subtracting the number of years of the female from the number of years of the male ($x_m - x_f$). $\beta 3$ is the estimator of the difference effect (of the interaction between meeting online and time period), X is a set of control variables and ε is the error term.

The control variables included are *male age* and *female age,* that represents the age of both partners in each observation. The control variables are chosen a priori, both income distribution and educational differences should vary with the relative age of the partners. The relatively older partner has had more time to pursue an education and career and should be more likely to earn more or have a higher degree. By controlling for the age of both partners, the estimated difference in the outcome dependent on age can be sorted out, which allows the focus to further lie on time period and means of meeting.

6.3 Wald tests

For each individual regression analysis, an additional test is executed. By using a Wald test, I can test whether the outcome of the online couples is statistically significantly different from the outcome of the offline couples in the 2017 sample only. By testing the following hypotheses:

$$H_0: \beta_1 + \beta_3 = 0$$
$$H_1: \beta_1 + \beta_3 \neq 0$$

For the 2009 sample, this question is already handled directly in the regression analysis by observing the estimate for the constant (β_0) and the online dummy (β_1). The Wald test tests whether the coefficient for the online variable (β_1) plus the coefficient for the interaction term (β_3) is statistically different from zero, as these variables represent the difference in outcome between the offline and online group in the 2017 sample.

6.4 Econometric considerations

The estimation model has the same construction as a normal difference-in-differences model. The difference-in-difference approach is usually leveraged when there is a treatment group, a control group and a causal effect of the treatment is to be determined (Lechner 2011). In that aspect, this study is different, instead of a treatment I have a *means of meeting*, and the market type is not meant to be interpreted as to have a causal effect on the outcome. Observations should potentially exhibit different outcomes depending on means of meeting, although, I am not claiming that means of meeting has a causal effect on sorting outcomes. The hypothesis is rather that the participants that choose (or are more successful in) the online market has certain characteristics, not that the market necessarily affect sorting outcomes. The interpretation of the estimators will therefore not imply causality but only detect potential patterns of differences across market type and time period. The model used should not be considered a true difference-in-differences approach rather a linear regression with an interaction.

The main motivation of using a model inspired of the difference-in-difference approach is that a potential *differential pattern* can be distinguished. In this case, the coefficient of the interaction term will represent the differential pattern of meeting online in the 2017 sample compared to the 2009 sample in the outcome variable. While using a linear regression without the interaction term would distinguish potential differences across the four groups, including the interaction term will yield more nuanced results. A reasonable alternative would be to conduct a set of t-tests. Although, an advantage of the chosen model compared to the t-tests is avoidance of the *multiple testing problem*, that arises when considering a set of statistical inferences simultaneously and results in an increase of the likelihood of obtaining erroneous inferences (Bender and Lange 2001).

All tests are executed in STATA, version 16.0. The multiple linear regression is executed with robust standard errors using the *robust option* in STATA. The point estimates of the coefficients are the same as they would be with a normal ordinary least square approach, although the standard errors take issues of heterogeneity and lack of normality into account. If an observation has any variable included in a regression analysis missing, then the observation is excluded from the analysis. The chosen level of statistical significance for interpretation of results in all analyses and tests are 5% ($p=0.05$).

7 Dataset

The *How Couples Meet and Stay Together* (HCMST) surveys have been conducted since 2009 by Stanford University, with principal investigator Michael Rosenfeld. Each couple is represented by the response of one partner. The respondents answer various questions about themselves, their partner and the relationship as such. The same respondents were studied in several waves of the survey and in 2017, a new sample with no overlap with the previous respondents was collected. I use the first wave (2009) and the fresh sample (2017) for this study. The data are considered to be nationally representative thanks to the sampling method (phone and Address Based Sampling, subjects without internet access were given access).

In this section, I present the exclusion criteria, that I follow to remove observations from the data; the coding of variables, that I execute to generate the educational and income distributional variables needed; and general descriptive statistics of the final sample, divided by year of collection.

7.1 Exclusion criteria

Non-qualified respondents (n=1109 excluded)

In 2009, all respondents without a current partner were considered disqualified by the survey providers and thus not allowed to proceed with the survey. In 2017, only respondents whom had never had a partner were judged disqualified: the respondents who currently did not have a partner were instead asked to answer questions about their previous relationship. Additionally, some other respondents were disqualified in the 2017 sample (n=9), because of not answering key questions. I choose to follow the original qualification criteria for both samples and thereby include respondents whom answered the questions about a past partner from the 2017 sample, even if that data is unavailable from 2009. I consider the variables used for this paper to be objective and I do not expect respondents to answer more falsely after a separation. Although these observations could potentially still be biased if couples with specific characteristics are more or less inclined to break up, then there would be an overrepresentation of these characteristics in the 2017 sample. The number of these respondents represented 132 of the 704 observation from the 2017 sample. To assure robust results, a robustness analysis is conducted by excluding these observations and repeating the full analysis, and thereafter controlling that the results remain unchanged in terms of statistical significance, sign and approximate size of estimators.

Lesbian, Gay and Bisexual oversampling (n=400 excluded)

In the 2017 survey, there was an oversampling of self-identified LBG respondents. As this inclusion is only present in the 2017 sample, I chose to exclude the entire oversampling. After this exclusion, LGB respondents are still present in both samples, although only with the natural proportion of the population. The reason for this exclusion is that I expect that the group might express different characteristics regarding intra-couple gender equality, and an inclusion in only one wave would potentially bias the data.

Same-sex couples (n=555 excluded)

Same-sex couples are present in both waves and excluded since gendered educational differences and income distribution between the genders are the aspects of interest.

Out-of-scope couples (n=3989 excluded)

I exclude all couples formed before 2009 in the 2017 sample and all couples formed before 2001 in the 2009 sample. The reason is that this exclusion is supported by the research question and econometric approach: as I study how matching outcomes of the two market types has changed over time, I want to eliminate the overlap in *year met* between the samples. Therefore, couples formed before 2009 are excluded from the 2017 sample, and to have a systematic approach, couples formed only in the last eight years are kept in the 2009 sample as well. This way, the 2017 sample represents matching outcomes generated after the 2009 sample was collected while the 2009 sample represent matching outcomes from the eight years leading up to 2009. Additionally, couples that were formed before the emergence of online markets are removed with a margin (first generation of online dating sites started in 1995, see section *2.2.1 history*).

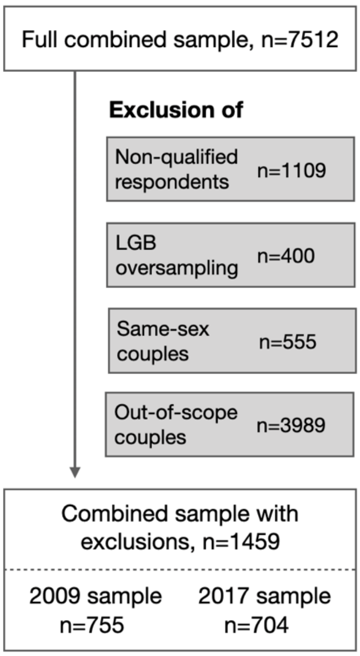

Figure 4. Overview of exclusions and sample sizes.
Notes. Illustration created by the author. Data source: 2009 and 2017 samples of HCMST datasets.

7.2 Coding of variables

Definition and specification of the dependent variables generated for this thesis are displayed in in table 1. Several existing variables in the HCMST dataset were used to generate the gender difference variables.

Table 1. Variable description

Name	Type of variable	Definition	Explanation of code written to generate variable	Existing variables used
Income equality (*equalincome*)	Dummy	Indication whether partners has about equal income or not	1 if the respondent answered: "We earned about the same amount" or, "My partner was not working for pay" and the respondents employment status was "not working". 0 in all other cases.	Q23 (who earned more income last year?) and ppwork (respondent's current employment status)
Male with higher income (*maleincome*)	Dummy	Indication whether male has higher income than the female	1 if the respondent answered: "I earned more" and was a male or, "My partner earned more" and was a female or, "My partner was not working for pay", was a male and was working. 0 in all other cases.	Q23 (who earned more income last year?), ppwork (respondent's current employment status) and ppgender (respondents' gender)
Female with higher income (*fenaleincome*)	Dummy	Indication whether female has higher income than the male	1 if the respondent answered: "I earned more" and was a female or, "My partner earned more" and was a male or, "My partner was not working for pay", was a female and was working. 0 in all other cases.	Q23 (who earned more income last year?), ppwork (respondents' current employment status) and ppgender (respondents' gender)
Educational difference (*educationdiff*)	Continuous (gap)	Difference between partners' years of educational attainment	Obtained by subtracting the female's years of education from the male's years of education. The difference becomes positive if the male has more education and negative if the female has more education. $(x_m - x_f)$	subject_yrsed (respondent's years of education), partner_yrsed (partners years of education) and ppgender (respondents' gender)
Maternal educational difference (*motherseducationdiff*)	Continuous (gap)	Difference between partners' mothers' years of educational attainment	Obtained by subtracting the female's mother's years of education from the male's mother's years of education. The difference becomes positive if the male's mother has more education and negative if the female's mother has more education. $(x_m - x_f)$	subject_mother_yrsed (respondent's mother's years of education), partner_mother_yrsed (partner's mother's years of education) and ppgender (respondents' gender)

Notes. The generation of the defined variables is only executed when the respondent has complete information in the variables used for construction (right most column). If the respondent lacks information for a variable, then the generated variable is labeled *missing*. If an observation has any variable included in a regression analysis missing, then the observation is excluded from the specific analysis. See appendix 1 for complete codebook. Data source: 2009 and 2017 HCMST datasets.

7.3 Summary statistics

The final dataset—constructed by the two waves and with systematic exclusions and generations described above—has the descriptive statistics displayed in table 2. The observations are separated horizontally by wave and descriptive statistics are displayed by variable of interest, a few additional relevant demographic variables are included. Each aspect is displayed in specification 1-5, respectively. A third column displays the *p*-value of a two-tailed t-test regarding the difference of means across samples. To show a more nuanced picture of educational background, full educational attainment and mother's educational attainment by gender is described, not only the gendered educational difference.

Table 2. Descriptive statistics by time period

	2009 sample	2017 sample	p-value
Number of observations N (%)	753 (0.52)	704 (0.48)	
1. Financial balance			
1a. Income equality N (%)	116 (0.16)	145 (0.21)	0.0082
1b. Male with higher income N (%)	439 (0.59)	364 (0.53)	0.0158
1c. Female with higher income N (%)	191 (0.26)	184 (0.27)	0.6825
2. Educational background			
2a. Mean years of male education (S.D)	13.85 (2.16)	13.73 (2.42)	0.3239
2b. Mean years of female education (S.D)	13.92 (2.11)	13.87 (2.27)	0.6644
2c. Mean years of male's mother's education (S.D)	12.92 (2.92)	13.01 (3.22)	0.5819
2d. Mean years of female's mother's education (S.D)	13.06 (2.79)	12.97 (2.88)	0.5283
3. Mean years since meeting (S.D)	3.75 (2.57)	3.77 (2.53)	0.8898
4. Mean age of male (S.D)	35.68 (13.58)	37.4 (14.68)	0.0211
5. Mean age of female (S.D)	33.35 (12.97)	35.44 (14.25)	0.0036

Notes. p-value for two-tailed t-test. Standard deviation or percentage in parenthesis. Data source: 2009 and 2017 HCMST datasets, displayed data includes the author's own coding.

There is a similar amount of observations from each wave (n=753 from 2009 and n=704 from 2017). Regarding income variables, a shift from higher income males to equal income partners has taken place during the time period, as the share of equal income couples has grown (from 16% to 21%, p=0.0082) and the share of couples with higher income males has decreased (from 59% to 53%, p=0.0158), while the share of couples with higher income females remains stable (26%-27%, no statistically significant difference). Only small differences in mean years of own and maternal educational attainment are found. The average age of males and females in the couples has increased (about 1.7 years for males, p=0.0211, and 2 years for females, p=0.0036). While the mean female age is constantly lower than the mean age of their male counterpart, the age gap seems to be is decreasing.

8 Results

The results section begins with a presentation of group level descriptive statistics (observations sorted by means of meeting and wave into four groups), subsequently, results from the regression analyses and associated Wald tests are presented.

8.1 Group level descriptive data

Descriptive statistics by wave and means of meeting is presented in table 3 and figures 5-9. In table 3, each aspect is displayed in specification 1-5, respectively. For both time periods, couples that met offline represents the largest group. There is a slightly higher representation of couples that met online in the 2017 sample (20% in the 2009 sample and 26% in the 2017 sample).

Regarding the income distribution, the point estimates for income equality (specification 1a. in table 3 and figure 5) are similar for both market types in 2009, while the point estimates seem more different in 2017. Although, no differences across market in any time period are statistically significant.

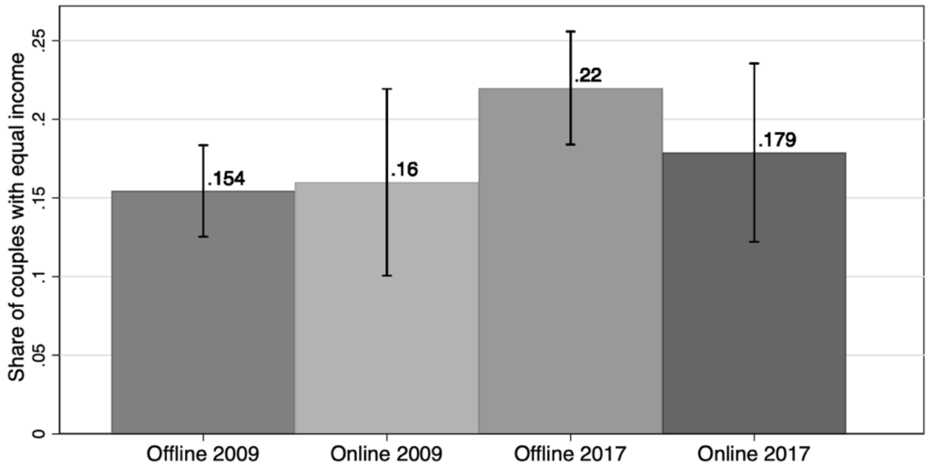

Figure 5. Share of couples with equal income across groups
Notes. Error bars indicate standard errors of the mean (SEM). Data source: 2009 and 2017 HCMST datasets, displayed data includes the author's own coding.

Exhibiting a male that earns more (specification 1b. and figure 6) was about equally common across both markets in the 2009 sample (59%). Even if the share of couples with this characteristic became less common with time, the difference in shares across markets seem to have increased to 2017, where the point estimate was a lower for online couples. Although, the difference across market type is not statistically significant.

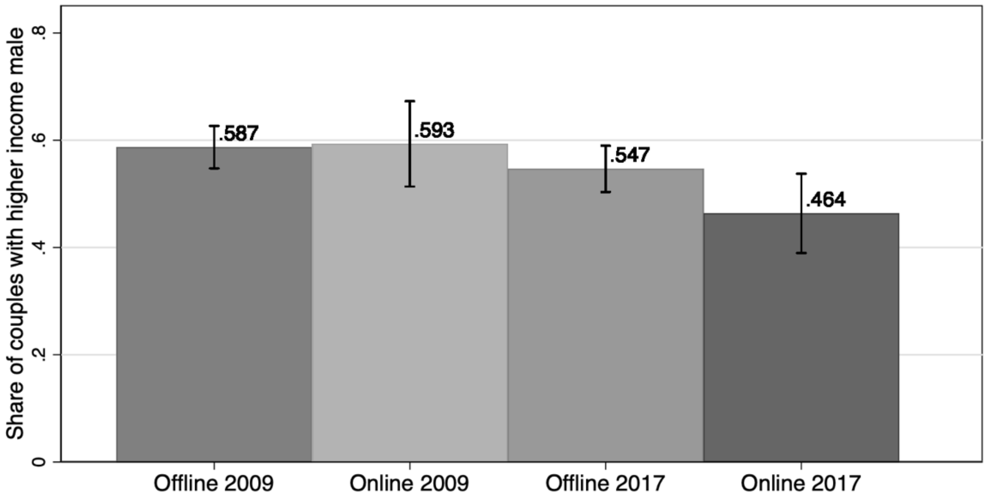

Figure 6. Share of couples with higher income male across groups
Notes. Error bars indicate standard errors of the mean (SEM). Data source: 2009 and 2017 HCMST datasets, displayed data includes the author's own coding.

A mirrored pattern can be found regarding the share of couples with a higher income female (specification 1c. and figure 7), In the 2009 sample, there was practically no difference across markets (around 25%) although in the 2017 sample, couples that met online displayed a statistically significantly higher share of couples with a higher income female (23% offline versus 36% online, p=0.0012).

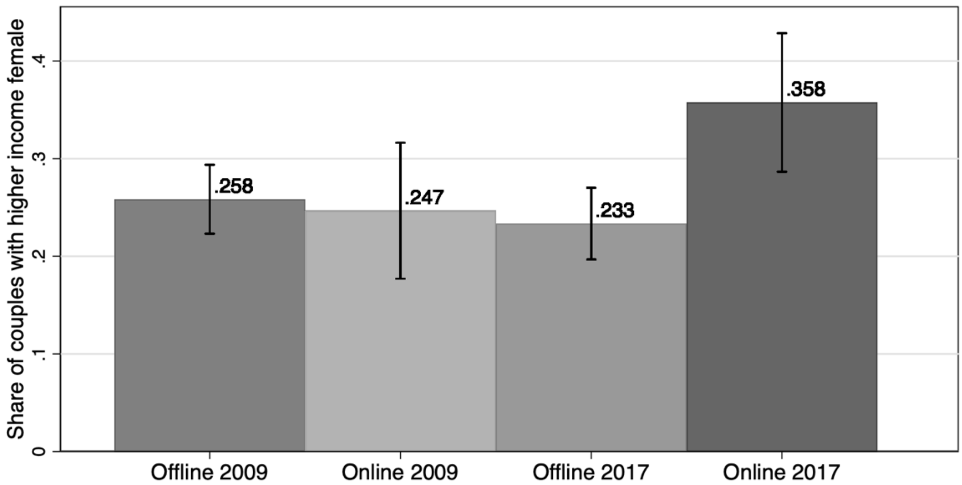

Figure 7. Share of couples with higher income female across groups
Notes. Error bars indicate standard errors of the mean (SEM). Data source: 2009 and 2017 HCMST datasets, displayed data includes the author's own coding.

Regarding education (specification 2a. and figure 8), the online group most often contain both males and females with higher mean years of education compared to the individuals that met offline (the difference is not statistically significant for males in the 2017 sample, p=0.0561). It could be because these individuals are on average older (Specification 4. and 5.). In the online groups, females have a higher point estimate for mean years of education compared to their male partners, although, the difference is not yet tested statistically.

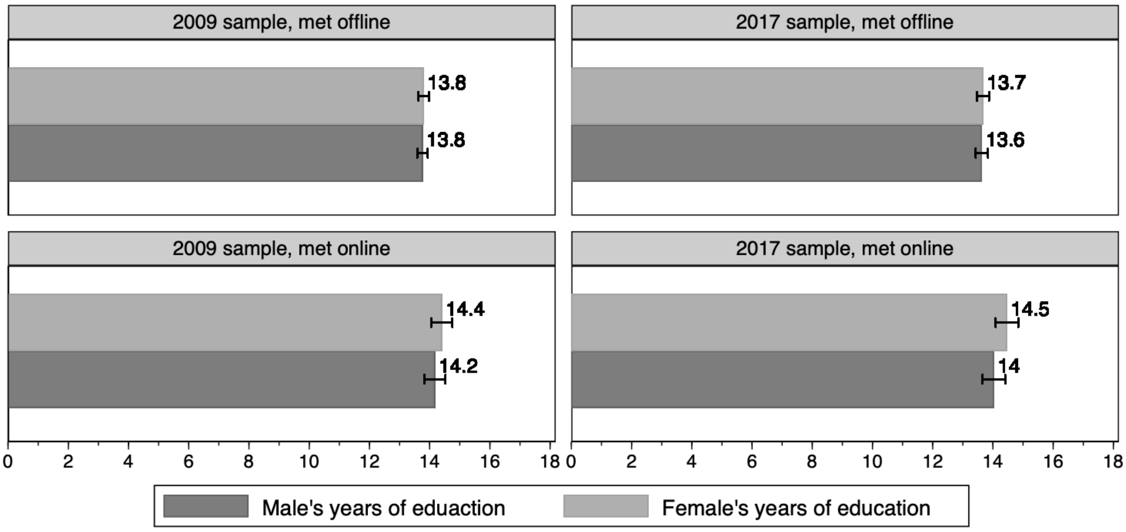

Figure 8. Average educational attainment separated by gender and group
Notes. Error bars indicate standard errors of the mean (SEM). Data source: 2009 and 2017 HCMST datasets, displayed data includes the author's own coding.

The point estimate for mean maternal education (2b. and figure 9) is very similar across markets for females in both time periods (13 years) but for males, the same is true only in the 2009 sample (13 years). In the 2017 sample however, men whose mothers has less education are slightly overrepresented in the online group (mean of 12.5 years online and 13.2 years offline, p=0.0288).

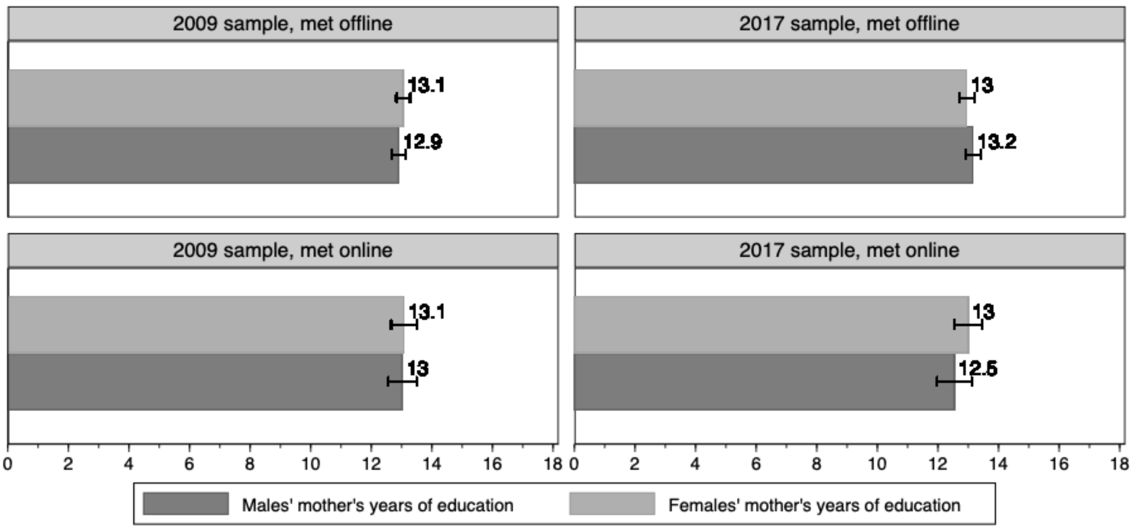

Figure 9. Average maternal educational attainment separated by gender and group
Notes. Error bars indicate standard errors of the mean (SEM). Data source: 2009 and 2017 HCMST datasets, displayed data includes the author's own coding.

Relationships initiated online are on average newer than those initiated offline in both time periods (specification 3.). On the other hand, both males and females are older in the online group compared to the offline group, a trait that seems to have increased over time.

Table 3. Descriptive statistics by studied groups

	2009 sample			2017 sample		
	Met offline	Met online	p-value	Met offline	Met online	p-value
Number of observations (%)	602(0.80)	151(0.20)		524(0.74)	180(0.26)	
1. Financial balance						
1a. Observations with income equality (%)	92 (0.15)	24 (0.16)	0.8650	113 (0.22)	32 (0.18)	0.2453
1b. Observations with higher income male (%)	350 (0.59)	89 (0.59)	0.8925	281 (0.55)	83 (0.46)	0.0556
1c. Observations with higher income female (%)	154 (0.26)	37 (0.25)	0.7691	120 (0.23)	64 (0.36)	0.0012
2. Educational background						
2a. Mean years of male education (S.D)	13.76 (2.15)	14.18 (2.19)	0.0363	13.63 (2.36)	14.03 (2.57)	0.0561
2b. Mean years of female education (S.D)	13.8 (2.12)	14.4 (1.99)	0.0016	13.67 (2.21)	14.46 (2.33)	0.0000
2c. Mean years of male's mother's education (S.D)	12.91 (2.9)	13.03 (2.91)	0.6552	13.16 (2.98)	12.55 (3.83)	0.0288
2d. Mean years of female's mother's education (S.D)	13.05 (2.82)	13.08 (2.66)	0.9091	12.96 (2.81)	13 (3.08)	0.8582
3. Mean years since first met (S.D)	3.87 (2.6)	3.3 (2.42)	0.0165	4 (2.49)	3.16 (2.57)	0.0002
4. Mean age of male (S.D)	35.1 (13.93)	37.96 (11.87)	0.0207	36.46 (14.46)	40.08 (15)	0.0043
5. Mean age of female (S.D)	32.82 (13.19)	35.45 (11.91)	0.0268	34.41 (13.94)	38.41 (14.75)	0.0012

Notes. *p*-value for two-tailed t-test. Standard deviation or percentage in parenthesis. Data source: 2009 and 2017 HCMST datasets, displayed data includes the author's own coding.

8.2 Estimation of differences in gender equality variables

Displayed in table 4 are the results of the linear regressions on intra-couple gender equality variables. Table 4 is horizontally separated by each dependent gender equality variable and on each variable, the regression analysis is presented together with the Wald test. Each variable is displayed in specification 1-5, respectively. The analysis of income variables (specification 1-3) used 1420 observations and the analysis of own and maternal educational background (specification 4. and 5.) used 1432 and 1387 observations, respectively. The differences in observations are due to missing variables for some observations (e.g. to be able to attribute *Maleincome* a value, the information under gender, employment status and income question has to be complete, see Table 1 for definitions). A set of robustness tests are conducted while keeping the number of observations constant at n=1380, which is the number of observations with all variables included in any analysis available. The robustness analyses are discussed under section 9.1 Robustness tests.

The income equality regression analysis (specification 1.), shows a statistically significant difference in time period, where the share of couples with equal income is larger in the 2017 sample (p=0.004), although no statistically significant pattern is found for market type (p=0.694). As the coefficient of the interaction is statistically insignificant (p=0.334), there does not seem to be a differential impact of meeting online in 2017 compared to 2009. The Wald test shows that no statistically significant difference across market type is found in the 2017 sample separately either (p=0.3301).

The analysis regarding the inequality of having a higher income male (specification 2.) does not exhibit statistically significant patterns across time (p=0.201), market (p=0.894) nor for the interaction (p=0.150). Although, the Wald test of difference across markets in the 2017 sample indicate a statistically significant difference (p=0.0265), where a lower share of couples that met online exhibit this type of inequality.

The pattern is more distinct regarding the inequality of having a female that earns more than the male (specification 3.), no statistically significant difference can be found across time nor market separately (p=0.243 and p=0.857, respectively). Although, the interaction term has a statistically significant positive coefficient (p=0.016). The statistically significant interaction indicates that there is a positive differential impact of meeting online in 2017 as compared to 2009 on the probability of having a female that earns more than the male. Where meeting online has a rather large and positive impact for the outcome in the 2017 sample compared to the 2009 sample, meeting online did not seem to have any distinguishable meaning for the outcome in the 2009 sample. The Wald test support the difference across markets in the 2017 sample (p=0.0014).

Moving on to the partners educational difference (specification 4.), where no statistically significant correlations are found at the chosen level of statistical significance in the regression analysis, nor is there any statistically distinguishable difference across markets in the 2017 sample detected by the Wald test (p=0.0517).

Likewise, regarding the maternal educational difference variable (specification 5.), no statistically significant patterns are detected at the chosen level of statistical significance. The regression for maternal education has relatively large standard errors, which indicate weak explaining abilities by the independent variables for the outcome in the proposed model.

The control variables, *male age* and *female age* proved to have statistically significant explanatory value in both income inequality regression analyses. The direction of impact regarding the genders respective age is as expected: The probability of having a higher income male increases with male age and decreases with female age, and the revered relationship is found for the probability of having a higher income female.

Table 4. OLS regression analyzing the difference of market and time period on outcome variables

Dependent variable	1. Equalincome	2. Maleincome	3. Femaleincome	4. Educationdiff	5. Motherseducationdiff
Regression analysis					
Time period	0.0683***	-0.0381	-0.0302	0.0254	0.3677
	(0.0237)	(0.0298)	(0.0259)	(0.1343)	(0.2003)
Met online	0.0133	-0.0060	-0.0072)	-0.1820	0.1595
	(0.0337)	(0.0454)	(0.0400)	(0.2133)	(0.3140)
Interaction	-0.0466	-0.0903	0.1369**	-0.2240	-0.7913
	(0.0482)	(0.0627)	(0.0568)	(0.2962)	(0.4659)
Male age	-0.0017	0.0089***	-0.0072***	0.0175	-0.0048
	(0.0013)	(0.0020)	(0.0017)	(0.0094)	(0.0158)
Female age	0.0002	-0.0067***	0.0065***	-0.0147	-0.0045
	(0.0013)	(0.0021)	(0.0018)	(0.0104)	(0.0168)
Constant	0.2064***	0.4977***	0.2959***	-0.1880	0.1650
	(0.0314)	(0.0387)	(0.0341)	(0.1707)	(0.2573)
Observations	1420	1420	1420	1432	1387
R-squared	0.0092	0.0239	0.0215	0.0065	0.0055
Wald test (*Met online + interaction=0)*					
F-value	F (1, 1414) = 0.95	F (1, 1414) = 4.93	F (1, 1414) = 10.31	F (1, 1426) = 3.79	F (1, 1381) = 3.33
p-value of F	0.3301	0.0265	0.0014	0.0517	0.0681

Notes. Robust standard errors in parenthesis. **, *** indicates statistical significance at the 5%, and 1% level, respectively. See Appendix 2 for complete STATA outputs. Description of dependent variables: equalincome, maleincome, femaleincome represent couples with the characteristic of having equal income, having a higher income male and having a higher income female, respectively. They are dummy outcome variables and an observation can have only one of the three income distributions. educationdiff and mothereducationdiff are differences in years of education and maternal education, respectively. The value in years of the female is subtracted from the value of years of the male $(x_m - x_f)$ to construct the variables. Data source: 2009 and 2017 HCMST datasets, displayed data includes the author's own coding.

9 Discussion

This section consists of a discussion regarding the results of robustness tests, general patterns discovered, implications for market design and society, strengths and weaknesses of the analysis and finally, I propose directions for future research in the area.

9.1 Robustness tests

Two sets of robustness tests of the complete analysis are conducted. The first, by excluding observations that has any of the variables used in any analysis labeled as missing. In these robustness tests I keep the number of observations constant across the five regression analyses (n=1380). The complete analysis is repeated and displayed in the first part of Appendix 3. In these robust tests, all statistically significant coefficients of the regression analyses remain statistically significant and have the same sign and approximate size. Additionally, all statistically significant Wald tests remained statistically significant.

The second set of robustness tests regards the exclusion criteria of *non-qualified respondents*. As described in section 7.1 Exclusion criteria, 132 unpartnered respondents in 2017 were allowed to take the survey and answer the questions about their past partner while all unpartnered respondents were disqualified in 2009. The inclusion could potentially bias the data if couples with specific characteristics are more inclined to break up, then there would be an overrepresentation of these characteristics in the 2017 sample. To assure that the results hold regardless of these respondents, a set of robustness tests are conducted by excluding these observations and repeating the full analysis. The full tests output is displayed in the latter part of Appendix 3. All statistically significant coefficients maintain statistical significance, sign and approximate size except for the *time period* coefficient of the income equality regression analysis, which lost its statistical significance. Additionally, all Wald tests maintain significance, although, the Wald test for the educational attainment analysis gained statistical significance by the exclusion, where it now seems as if couples formed online in 2017 further express couples with females of relatively higher education compared to 2017 offline couples. To conclude, all results are robust in the discussed aspects except for the finding regarding the growth income equality over time.

9.2 Patterns for intra-couple gender (in)equality across market type

To summarize the main patterns found: Firstly, the trait of having a higher income male is more uncommon for online couples in 2017 compared to offline couples in 2017. Secondly, for the trait of having a higher income female, there is a distinct differential trend over time, where 2017 online couples exhibit the trait further than 2009 online couples. There is also a clear difference in the share of couples with a higher income female between market type in 2017, where online couples express the trait further.

The findings partly support the hypothesis regarding differences across markets, only in the 2017 sample does online couples exhibit relatively financially stronger females, weaker males and no more equality compared to the offline couples. No statistically significant differences are found in the 2009 sample regarding income distribution. The analyses of educational background find no statistically significant results. Additionally, the findings also partly support the hypothesized development over time. As hypothesized, these differences across market would have grown over time, and while no statistically significant differences across markets for any trait are found in the 2009 data, both income inequalities (male and female with higher income) differ across markets in 2017. This result indicates that differences across markets has become more distinguishable over time. It is interesting that no differences at all were found across markets in 2009, as the socio-demographic imbalance was present during 2001-2009 as well, although smaller in size. The findings are consistent with the proposed framework regarding differences of the online market—where participants are in abundance but often incapable of efficient evaluation, in a national setting of socio-demographic gender imbalances. The results partly support the suggestion of connecting micro-level market structures to macro-level socio-demographics to predict sorting outcomes. The macro-level imbalance is found to different extents in couples generated by the two markets, where the man deficit shines through the online market further. It makes sense that participants join the thick online market when they look for a kind of partner which is scarce in society, although, it is not possible for all participants to find a partner of their own/desired social class. Arguably, the online market is a good tool for singles to find a partner of a different background since they could be hard to find in their natural environment.

It is interesting that differences in the income inequality variables hold across markets in the 2017 sample while controlling for the age of both partners. Couples that met online then display further financially strong women and weak men compared to couples that met offline regardless of their relative age.

To build on previous research that identified further exogamy in couples that met online (Lee 2016; Potarca 2017; Thomas 2020), I can now add findings regarding the direction of exogamy across market type, where a clear difference in the direction of the financial inequality is detected in 2017. Online couples more often exhibit an income distribution that favors the female and more seldom exhibit an income distribution that favors the male compared to offline couples.

9.3 Implications for society

The societal implications of the findings are that the online market currently has a crucial function in society, where otherwise thin market participants can find a partner. The online market could reduce loneliness but also affect social stratification since it supports union between individuals of mixed social classes. In this sense, the online market could increase both intra-generational and intergenerational mobility in society and thereby reduce inequality. Also, if the online markets match singles that would otherwise choose to stay single, then the online markets could affect the number of households and births in the US. The results also shed light on the role of online dating in the challenge of imbalanced socio-demographics across genders.

9.4 Implications for market design

If my hypothesis is correct, that the current online market is thick, yet congested and unsafe—and thereby mainly serves participants of otherwise thin markets, then there are specific design implications associated with the result. The online dating sites should oversee the mechanisms and environment provided and consider re-designing the market structure. The marketplace could be improved to provide a good alternative for all kinds of participants, not only those unable to find alternatives offline. To do this, the online market structure needs to be designed to decrease congestion and provide safety.

One design implication is to increase transparency regarding competition, in order to reduce biased participant behavior due to thickness found by Fong (2019). For example, the site can simply display some distributional information to each user: the number of other participants from the same side with similar age (the competition) and the number of potential partners with similar age that they have to compete for. Another design implication is to increase consequences of approaching potential partners that is 'out of one's league', as studied by Kreager et al. (2014), to prevent congestion. This could be done by limiting the number of approaches allowed in a day or by displaying embarrassing information of potential previous rejections on each user's profile[17].

[17] A feature that is arguably controversial, although, might accommodate participants to act more similar to the *matching hypothesis* (proposed by Walster et al. (1966)) by increasing the cost of rejection and thereby decreasing congestion.

This way, the online structure would potentially accommodate participants to behave more like they do offline, for example in a bar. Additionally, efforts aimed at providing a safe and truthful environment is needed. By for example, identifying users with verified information before allowing them into the marketplace, this action could reduce informational asymmetries, identify criminals and reduce congestion by improving the ability to evaluate options.

9.5 Strengths and weaknesses of analysis

A key strength lies in the ability to leverage two sets of nationally representative data, with which I am further able to draw conclusions for general patterns of sorting in society. This is a valuable property of this paper. Having the opportunity of using two datasets from the same provider from two points in time ensures a consistently high quality of the data, and the fact that the two surveys were conducted eight years apart enabled me to collect valuable information.

The first noticeable weakness is that the data is quite thin for some aspects. While the original HCMST datasets were relatively large, I used them to study a narrow aspect of courtship (online dating). Additionally, while using relatively small subsets of observations I choose to study characteristics that were uncommon (e.g. only 24 observations of the online group from 2009 exhibited income equality). These combinations in study design weakened the statistical power of the analysis. Although, with the sparsity of available data in this area there is no reasonable alternative data source that could link matching market and sorting in society. The thinness of data on couples that met online incapacitated me from sorting these couples into more granular groups based on which online or offline market that was used in couple formation. A categorization that would have been highly valuable in this study.

A second weakness that I recognize regards the construction of income dummy variables. As summarized in table 1, they build on up to three existing variables in the dataset: the answer of the income question, the respondent's employment status and respondent's gender. If there were information regarding the employment status of the partner too, I would have been able to detect more precise income balances. It would also be favorable if information regarding the size of the potential income difference would be present in the data.

Finally, statistically significant patterns are not found for either educational difference analysis while controlling for age, which is also an interesting observation. This finding could indicate that agents using the online market are not essentially different from other individuals, just at a later point in life. There is also a possibility that that the analyses of dummy variables were more robust than the analysis of continuous gaps, and that potential differences across education could not be detected by design.

9.6 Future research

After studying the topic and conducting this analysis, I would like to bring forward four tracks regarding future research in this area.

Firstly, similar studies as this paper, but with more data could further validate the role of online dating in intra-couple gender equality. This kind of approach largely contributes to the areas of modern intra-generational mobility and social stratification, as emerging matching markets which seem to accommodate fundamentally different sorting patterns compared to traditional markets are studied.

Secondly, studies focused more generally on the role of online dating in society and its design properties would be useful to further understand self-selection of participants into different markets and its generation of sorting patterns. The importance of this area lies in understanding how and why online markets exhibit different sorting outcomes compared to offline markets, the research could provide important implications for market design in dating.

A third consideration is causality. I was not able to isolate the part of the patterns generated by self-selection into market types and potential causal impact generated by market type. To separate the two aspects would contribute to understanding how technology can assist our mate selection and how technology can fundamentally change the types of families we form.

Finally, an arguably important aspect is the commerciality of the emerging online markets. The business models of online dating services generate revenue from singles that use the service—representing a fundamental misalignment between users and corporation.[18] The user's goal is assumingly to find a partner and leave the market, while the corporation's goal is arguably to maintain a customer base. With the current revenue models, there is an incentive for corporations to provide a matching market structure that is unfavorable for efficient and stable match making. I consider it important to capture this aspect in future studies and to sort out whether online dating

[18] As presented under section 2.2 Dating markets, revenue models include running ads for users and paid memberships.

markets are at risk regarding intentional inefficiency. The results could provide crucial implications regarding the actors of the industry and might even encourage governmental efforts to provide a functional service for their people, in form of a public market or regulation. Perhaps, in a future society, a match making institution would be as needed, appreciated and taken for granted as education.

10 Conclusions

I propose a model framework that connect characteristics of market structures to the underlying macro-environment, in form of the socio-demographic gender balance. The online markets are considered inefficient but thick after critical evaluation and are thereby judged to mainly provide a valuable solution for participants of otherwise thin markets. In the US case, I identify thin market participants as highly educated women and men of less education. I propose that these participants are enabled to further find each other in online market structures and that couples that are formed online should further exhibit traits of financially and educationally strong females relative to their male partners, and that the internal inequality should have grown over time, together with the US socio-demographic imbalance. My hypotheses are partly confirmed by empirically studying two nationally representative datasets on US adult couples. I find that couples formed online from 2009 to 2017 exhibit further financially strong women in relation their male partner, compared to couples that were formed offline during the time period. Additionally, there is a distinct differential trend over time, where couples that formed online from 2009 to 2017 exhibit the trait further than couples formed online from 2001 to 2009. The findings add to previous research regarding exogamy in online dating and provide an important set of implications for both market design in online dating and social stratification in society.

11 Bibliography

Written sources

ABRAMOVA, O., BAUMANN, A., KRASNOVA, H. and BUXMANN, P., 2016. Gender Differences in Online Dating: What Do We Know So Far? A Systematic Literature Review. *2016 49th Hawaii International Conference on System Sciences (HICSS)*, USA: IEEE Computer Society. pp. 3858-3867.

BECKER, G.S., 1973. A Theory of Marriage: Part I. *Journal of Political Economy,* **81**(4), pp. 813-846.

BECKER, G.S., 1974. A Theory of Marriage: Part II. *Journal of Political Economy,* **82**(2), pp. 11-26.

BENDER, R. and LANGE, S., 2001. Adjusting for Multiple Testing—When and How?. *Journal of Clinical Epidemiology,* **54**(4), pp. 343-349.

BIRGER, J., 2015-last update. For the College-Educated, There Is a Man Deficit in the U.S. Available: https://www.nytimes.com/roomfordebate/2015/07/06/has-being-single-in-america-changed/for-the-college-educated-there-is-a-man-deficit-in-the-us [May 4, 2020].

DROUIN, M., MILLER, D., WEHLE, S.M.J., and HERNANDEZ, E., 2016. Why Do People Lie Online? "Because Everyone Lies on the Internet". *Computers in Human Behavior,* **64**, pp. 134-142.

EIKA, L., MOGSTAD, M. and ZAFAR, B., 2014. Educational Assortative Mating and Household Income Inequality. *Journal of Political Economy*, **127**(6), pp. 2795-2835.

FINKEL, E.J., EASTWICK, P.W., KARNEY, B.R., REIS, H.T. and SPRECHER, S., 2012. Online Dating: A Critical Analysis From the Perspective of Psychological Science. *Psychological Science in the Public Interest,* **13**(1), pp. 3-66.

FRISHBERG, H., 2019-last update. Broke Men Are Hurting American Women's Marriage Prospects. Available: https://nypost.com/2019/09/06/broke-men-are-hurting-american-womens-marriage-prospects/ [May 4, 2020].

FONG, J., 2019. Search, Selectivity, and Market Thickness in Two-Sided Markets (Working Paper).

GALE, D. and SHAPLEY, L.S., 1962. College Admissions and the Stability of Marriage. *The American Mathematical Monthly,* **69**(1), pp. 9-15.

GELLES, D., 2011-last update, Inside Match.com, It's All About the Algorithm. Available: http://www.slate.com/id/2300430/ [May 4, 2020].

GUTTENTAG, M. and SECORD, P.F., 1983. *Too Many Women? The Sex Ratio Question*. Beverly Hills: SAGE Publications.

GREENWOOD, J., GUNER, N., KOCHARKOV, G., and SANTOS, C., 2014. Marry Your Like: Assortative Mating and Income Inequality. *American Economic Review*, **104** (5), pp. 348-53.

HANCOCK, J., TOMA, C. and ELLISON, N., 2007. The Truth About Lying in Online Dating Profiles. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2007),* pp. 449-452.

HITSCH, G.J., HORTAÇSU, A. and ARIELY, D., 2010. Matching and Sorting in Online Dating. *The American Economic Review,* **100**(1), pp. 130-163.

JACKSON, M.I., KIERNAN, K., MCLANAHAN, S. and MOFFITT, S.L., 2017. Maternal Education, Changing Family Circumstances, and Children's Skill Development in the United States and UK. *The Annals of the American Academy of Political and Social Science*, **674**(1), pp. 59-84.

KREAGER, D.A., CAVANAGH, S.E., YEN, J. and YU, M., 2014. "Where Have All the Good Men Gone?" Gendered Interactions in Online Dating. *Journal of Marriage and Family,* **76**(2), pp. 387-410.

LECHNER, M., 2011. The Estimation of Causal Effects by Difference-in-Difference Methods, *Foundations and Trends(R) in Econometrics,* **4**(3), pp 165-224.

LEE, S. and NIEDERLE, M., 2015. Propose With a Rose? Signaling in Internet Dating Markets. *Experimental Economics,* **18**(4), pp. 731-755.

LEE, S., 2016. Effect of Online Dating on Assortative Mating: Evidence from South Korea. *Journal of Applied Econometrics,* **31**(6), pp. 1120-1139.

LICHTER, D.T., PRICE, J.P. and SWIGERT, J.M., 2020. Mismatches in the Marriage Market. *Journal of Marriage and Family,* **82**(2), pp. 796-809.

MARKOWITZ, D. and HANCOCK, J., 2013. Deception in Mobile Dating Conversations. *Journal of Communication,* **68**(3), pp. 547–569.

MELTZER, M., 2016, Online Dating: Match Me If You Can. Available: https://www.consumerreports.org/dating-relationships/online-dating-guide-match-me-if-you-can/ [May 4, 2020].

PEW RESEARCH CENTER, 2020. The Virtues and Downsides of Online Dating. Available: https://www.pewresearch.org/internet/2020/02/06/the-virtues-and-downsides-of-online-dating/ [May 4, 2020].

POTARCA, G., 2017. Does the Internet Affect Assortative Mating? Evidence From the U.S. and Germany. *Social Science Research,* **61**, pp. 278-297.

RAYMO, J.M. and IWASAWA, M., 2005. Marriage Market Mismatches in Japan: An Alternative View of the Relationship between Women's Education and Marriage. *American Sociological Review,* **70**(5), pp. 801-822.

ROSENFELD, M.J. and THOMAS, R. J., 2012. Searching for a Mate: The Rise of the Internet as a Social Intermediary. *American Sociological Review,* **77**(4), pp. 523-547.

ROSENFIELD, M., 2017. Marriage, Choice, and Couplehood in the Age of the Internet. *Sociological Science,* **4**(20), pp. 490-510.

ROSENFELD, M., THOMAS, R. and HAUSEN, S., 2019. Disintermediating Your Friends: How Online Dating in the United States Displaces Other Ways of Meeting. *Proceedings of the National Academy of Sciences of the United States of America,* **116**(36), pp. 17753-17758.

ROTH, A.E., 2002. The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics. *Econometrica,* **70**(4), pp. 1341-1378.

ROTH, A.E. and SONMEZ, T., 2004. Kidney Exchange. *The Quarterly Journal of Economics,* **119**(2), pp. 457-488.

ROTH, A.E., 2008. What Have We Learned from Market Design?. *Economic Journal,* **118**(527), pp. 285-310.

ROTH, A.E., 2015. *Who Gets What—and Why: The New Economics of Matchmaking and Market Design.* Boston: Houghton Mifflin Harcourt.

THE FEDERAL BUREAU OF INVESTIGATION (FBI) INTERNET CRIME COMPLAINT CENTER (IC3), 2020. *2019 Internet Crime Report.* Available: https://pdf.ic3.gov/2019_IC3Report.pdf [June 2, 2020].

THOMAS, R.J., 2020. Online Exogamy Reconsidered: Estimating the Internet's Effects on Racial, Educational, Religious, Political and Age Assortative Mating. *Social Forces,* **98**(3), pp. 1257–1286.

TINDER, March 15, 2019-last update, Powering Tinder — The Method Behind Our Matching. Available: https://blog.gotinder.com/powering-tinder-r-the-method-behind-our-matching/ [May 4, 2020].

WALSTER, E., ARONSON, V., ABRAHAMS, D. and ROTTMAN, L., 1966. Importance of Physical Attractiveness in Dating Behavior. *Journal of Personality and Social Psychology,* **4**(5), pp. 508-516.

WANG, W. and PARKER, K., 2014. Record Share of Americans Have Never Married: As Values, Economics and Gender Patterns Change. Washington, D.C.: Pew Research Center's Social & Demographic Trends project. Available: https://www.pewsocialtrends.org/wp-content/uploads/sites/3/2014/09/2014-09-24_Never-Married-Americans.pdf [June 2, 2020].

Data

NATIONAL CENTER FOR EDUCATION STATISTICS (NCES), 2019-last update, Number of Bachelor's Degrees Earned in the United States From 1949/50 to 2028/29, by Gender (in 1,000s). Available: https://www-statista-com.ez.hhs.se/statistics/185157/number-of-bachelor-degrees-by-gender-since-1950/
[May 4, 2020].

NATIONAL CENTER FOR EDUCATION STATISTICS (NCES), 2019-last update, Number of Masters's Degrees Earned in the United States From 1949/50 to 2028/29, by Gender (in 1,000s). Available: https://www-statista-com.ez.hhs.se/statistics/185160/number-of-masters-degrees-by-gender-since-1950/
[May 4, 2020].

ROSENFELD, M.J., THOMAS, R.J. and FALCON, M., 2018. How Couples Meet and Stay Together, Waves 1, 2, and 3: Public version 3.04, plus wave 4 supplement version 1.02 and wave 5 supplement version 1.0 and wave 6 supplement ver 1.0 [Computer files]. Stanford, CA: Stanford University Libraries.

ROSENFELD, M.J., THOMAS, R.J. and HAUSEN, S., 2019. How Couples Meet and Stay Together 2017 fresh sample. [Computer files]. Stanford, CA: Stanford University Libraries.

Appendix 1: Complete codebook

```
/* generate sorting variables */
generate wave = .
replace wave = 1 if CaseID > 0.5
replace wave = 0 if CaseID == .

generate same_sex = .
replace same_sex = 1 if same_sex_couple == 1
replace same_sex = 1 if w6_same_sex_couple  == 1
replace same_sex = 0 if same_sex_couple == 0
replace same_sex = 0 if w6_same_sex_couple == 0

generate met_online = .
replace met_online = 1 if hcm2017q24_met_online == 1
replace met_online = 0 if hcm2017q24_met_online == 0
replace met_online = 1 if either_internet_adjusted == 1
replace met_online = 0 if either_internet_adjusted == 0
replace met_online = 0 if either_internet_adjusted == -1

generate first_met = .
replace first_met = (ppage - age_when_met) if wave == 1 & !missing(ppage, age_when_met)
replace first_met = (ppage - q21a) if wave == 0 & !missing(ppage, q21a)

generate year_met = (2009 - first_met) if wave == 0 & !missing(first_met)
replace year_met = (2017 - first_met) if wave == 1 & !missing(first_met)

generate exclude = 1 if year_met < 2009 & wave == 1 & !missing(year_met)
replace exclude = 0 if year_met >= 2009 & wave == 1 & !missing(year_met)
replace exclude = 1 if year_met < 2001 & wave == 0 & !missing(year_met)
replace exclude = 0 if year_met >= 2001 & wave == 0 & !missing(year_met)

generate missing = 1 if missing(inc_eq, m_inc_ineq, f_inc_ineq, ed_gap, m_ed_gap)

generate group = 1 if wave == 0 & met_online == 0
replace group = 2 if wave == 0 & met_online == 1
replace group = 3 if wave == 1 & met_online == 0
replace group = 4 if wave == 1 & met_online == 1

/* generate descriptives */
generate p_age = w6_q9 if wave == 1
replace p_age = q9 if wave == 0
replace p_age = .  if p_age == -1
```

```
generate male_age = ppage if ppgender == 1 & !missing(ppage, ppgender)
replace male_age = p_age if ppgender == 2 & !missing(p_age, ppgender)
generate female_age = ppage if ppgender == 2 & !missing(ppage, ppgender)
replace female_age = p_age if ppgender == 1 & !missing(p_age, ppgender)

replace subject_yrsed = respondent_yrsed if wave == 0
replace subject_mother_yrsed = respondent_mom_yrsed if wave == 0
replace partner_mother_yrsed = partner_mom_yrsed if wave == 0

generate male_ed = subject_yrsed if ppgender == 1 & !missing(subject_yrsed, ppgender)
replace male_ed = partner_yrsed if ppgender == 2 & !missing(partner_yrsed, ppgender)

generate female_ed = subject_yrsed if ppgender == 2 & !missing(subject_yrsed, ppgender)
replace female_ed = partner_yrsed if ppgender == 1 & !missing(partner_yrsed, ppgender)

generate male_m_ed  = subject_mother_yrsed if ppgender == 1 & !missing(subject_mother_yrsed, ppgender)
replace male_m_ed  = partner_mother_yrsed if ppgender == 2 & !missing(partner_mother_yrsed, ppgender)

generate female_m_ed  = subject_mother_yrsed if ppgender == 2 & !missing(subject_mother_yrsed, ppgender)
replace female_m_ed  = partner_mother_yrsed if ppgender == 1 & !missing(partner_mother_yrsed, ppgender)

/* generate gaps */
generate ed_gap = male_ed - female_ed if !missing(male_ed, female_ed)
replace ed_gap = . if female_ed == -1
replace ed_gap = . if male_ed == -1

generate m_ed_gap = male_m_ed - female_m_ed if !missing(male_m_ed, female_m_ed)
replace m_ed_gap = . if male_m_ed == -1
replace m_ed_gap = . if female_m_ed == -1

/* generate dummies */
generate income_q = q23 if wave == 0 & !missing(q23)
replace income_q = w6_q23 if wave == 1 & !missing(w6_q23)
replace income_q = . if income_q == -1

generate inc_eq = 0
replace inc_eq = 1 if income_q == 2
replace inc_eq = 1 if income_q == 4 & ppwork > 2.5
replace inc_eq = . if income_q == -1
replace inc_eq = . if income_q == .

generate m_inc_ineq = 0
replace m_inc_ineq = 1 if income_q == 1 & ppgender == 1
replace m_inc_ineq = 1 if income_q == 3  & ppgender == 2
```

Appendix 2: Full test outputs for table 4

```
      name:  <unnamed>
       log:  /Users/lovisaqvarner/Downloads/main_analysis.smcl
  log type:  smcl
 opened on:  30 Apr 2020, 14:29:17

1 . reg inc_eq i.wave##i.met_online  male_age female_age, robust

Linear regression                               Number of obs     =      1,420
                                                F(5, 1414)        =       2.32
                                                Prob > F          =     0.0412
                                                R-squared         =     0.0092
                                                Root MSE          =     .38346
```

inc_eq	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
1.wave	.0682686	.0236964	2.88	0.004	.0217847	.1147526
1.met_online	.0132653	.0336882	0.39	0.694	-.0528189	.0793494
wave#met_online						
1 1	-.0465589	.0482086	-0.97	0.334	-.141127	.0480093
male_age	-.0017261	.0013058	-1.32	0.186	-.0042875	.0008354
female_age	.0001996	.0013436	0.15	0.882	-.002436	.0028353
_cons	.2064021	.0314094	6.57	0.000	.1447881	.2680161

```
2 . test (1.met_online + 1.wave#1.met_online = 0)

 ( 1)  1.met_online + 1.wave#1.met_online = 0

       F(  1,  1414) =    0.95
            Prob > F =    0.3301

3 . reg m_inc_ineq i.wave##i.met_online male_age female_age, robust

Linear regression                               Number of obs     =      1,420
                                                F(5, 1414)        =       6.75
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0239
                                                Root MSE          =     .49147
```

m_inc_ineq	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
1.wave	-.0381008	.029793	-1.28	0.201	-.0965439	.0203424
1.met_online	-.0060459	.0453708	-0.13	0.894	-.0950471	.0829554
wave#met_online						
1 1	-.0903429	.0626841	-1.44	0.150	-.2133067	.0326209
male_age	.0088886	.00197	4.51	0.000	.0050241	.012753
female_age	-.0067386	.0021188	-3.18	0.002	-.010895	-.0025823
_cons	.4976965	.0387268	12.85	0.000	.4217284	.5736647

```
4 . test (1.met_online + 1.wave#1.met_online = 0)

 ( 1)  1.met_online + 1.wave#1.met_online = 0

       F(  1,  1414) =    4.93
            Prob > F =    0.0265

5 . reg f_inc_ineq i.wave##i.met_online  male_age female_age, robust

Linear regression                               Number of obs     =      1,420
                                                F(5, 1414)        =       5.76
                                                Prob > F          =     0.0000
```

```
                                                R-squared         =     0.0215
                                                Root MSE          =     .43473
```

f_inc_ineq	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
1.wave	-.0301679	.0258557	-1.17	0.243	-.0808874	.0205517
1.met_online	-.0072194	.0400185	-0.18	0.857	-.0857214	.0712826
wave#met_online						
1 1	.1369017	.0567756	2.41	0.016	.0255283	.2482752
male_age	-.0071625	.0016794	-4.26	0.000	-.0104569	-.0038681
female_age	.006539	.0018077	3.62	0.000	.002993	.010085
_cons	.2959013	.0340646	8.69	0.000	.2290788	.3627239

```
6 . test (1.met_online + 1.wave#1.met_online = 0)

 ( 1)  1.met_online + 1.wave#1.met_online = 0

       F(  1,  1414) =   10.31
            Prob > F =    0.0014

7 . reg ed_gap i.wave##i.met_online male_age female_age, robust

Linear regression                               Number of obs     =      1,432
                                                F(5, 1426)        =       1.53
                                                Prob > F          =     0.1775
                                                R-squared         =     0.0065
                                                Root MSE          =      2.265
```

ed_gap	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
1.wave	.0254255	.1342541	0.19	0.850	-.2379312	.2887821
1.met_online	-.1820115	.2132724	-0.85	0.394	-.6003728	.2363499
wave#met_online						
1 1	-.2239807	.2961987	-0.76	0.450	-.8050127	.3570513
male_age	.017458	.0094065	1.86	0.064	-.0009941	.03591
female_age	-.0147396	.010384	-1.42	0.156	-.0351093	.00563
_cons	-.1879893	.1706813	-1.10	0.271	-.5228026	.1468241

```
8 . test (1.met_online + 1.wave#1.met_online = 0)

 ( 1)  1.met_online + 1.wave#1.met_online = 0

       F(  1,  1426) =    3.79
            Prob > F =    0.0517

9 . reg m_ed_gap i.wave##i.met_online male_age female_age, robust

Linear regression                               Number of obs     =      1,387
                                                F(5, 1381)        =       1.43
                                                Prob > F          =     0.2088
                                                R-squared         =     0.0055
                                                Root MSE          =     3.3832
```

m_ed_gap	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
1.wave	.367741	.2002802	1.84	0.067	-.0251453	.7606273
1.met_online	.1595005	.3139578	0.51	0.612	-.4563854	.7753863
wave#met_online						

```
            1 1  |  -.7913213   .4659135    -1.70   0.090    -1.705296    .1226533
                 |
        male_age |  -.0047794   .0158072    -0.30   0.762    -.0357882    .0262294
      female_age |  -.0044725   .0167822    -0.27   0.790    -.0373939    .0284489
           _cons |   .1649839   .2573421     0.64   0.522    -.3398399    .6698076
------------------------------------------------------------------------------------

10 . test (1.met_online + 1.wave#1.met_online = 0)

 ( 1)  1.met_online + 1.wave#1.met_online = 0

       F(  1,  1381) =    3.33
            Prob > F =    0.0681

11 . log close
      name:  <unnamed>
       log:  /Users/lovisaqvarner/Downloads/main_analysis.smcl
  log type:  smcl
 closed on:  30 Apr 2020, 14:29:24
```

Appendix 3: Full test outputs for robustness tests

Exclusion of observations with incomplete information

```
      name:  <unnamed>
       log:  /Users/lovisaqvarner/Desktop/Main_robust.smcl
  log type:  smcl
 opened on:  30 Apr 2020, 14:38:05

1 . reg inc_eq i.wave##i.met_online male_age female_age if missing ==., robust

Linear regression                               Number of obs     =      1,380
                                                F(5, 1374)        =       2.74
                                                Prob > F          =     0.0182
                                                R-squared         =     0.0107
                                                Root MSE          =     .38089

-------------------------------------------------------------------------------
               |               Robust
        inc_eq |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------+---------------------------------------------------------------
        1.wave |   .0713575   .0239245     2.98   0.003      .024425    .1182901
  1.met_online |   .0131878   .0335498     0.39   0.694    -.0526266    .0790022
               |
wave#met_online |
          1 1  |   -.047868   .0485507    -0.99   0.324    -.1431095    .0473735
               |
      male_age |  -.0021637   .0013017    -1.66   0.097    -.0047172    .0003898
    female_age |   .0004678    .001352     0.35   0.729    -.0021844    .0031199
         _cons |   .2092114   .0315333     6.63   0.000     .1473528      .27107
-------------------------------------------------------------------------------

2 . test (1.met_online + 1.wave#1.met_online = 0)

 ( 1)  1.met_online + 1.wave#1.met_online = 0

       F(  1,  1374) =    1.00
            Prob > F =    0.3178

3 . reg m_inc_ineq i.wave##i.met_online male_age female_age if missing ==., robust

Linear regression                               Number of obs     =      1,380
                                                F(5, 1374)        =       7.23
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0261
                                                Root MSE          =     .49057

-------------------------------------------------------------------------------
               |               Robust
    m_inc_ineq |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------+---------------------------------------------------------------
        1.wave |  -.0365441   .0301699    -1.21   0.226    -.0957281    .0226399
  1.met_online |  -.0095319   .0457428    -0.21   0.835    -.0992651    .0802013
               |
wave#met_online |
          1 1  |  -.0957489   .0635211    -1.51   0.132    -.2203578      .02886
               |
      male_age |   .0093862   .0020174     4.65   0.000     .0054287    .0133437
    female_age |  -.0070118   .0021678    -3.23   0.001    -.0112642   -.0027593
         _cons |   .4925793   .0390979    12.60   0.000     .4158811    .5692774
-------------------------------------------------------------------------------

4 . test (1.met_online + 1.wave#1.met_online = 0)

 ( 1)  1.met_online + 1.wave#1.met_online = 0

       F(  1,  1374) =    5.67
            Prob > F =    0.0174

5 . reg f_inc_ineq i.wave##i.met_online male_age female_age if missing ==., robust

Linear regression                               Number of obs     =      1,380
                                                F(5, 1374)        =       5.84
                                                Prob > F          =     0.0000
```

```
                                        R-squared         =     0.0227
                                        Root MSE          =     .43426
```

f_inc_ineq	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
1.wave	-.0348134	.0261632	-1.33	0.184	-.0861377	.0165108
1.met_online	-.0036559	.0404884	-0.09	0.928	-.0830816	.0757698
wave#met_online						
1 1	.1436169	.0577074	2.49	0.013	.0304127	.256821
male_age	-.0072224	.0017314	-4.17	0.000	-.0106189	-.003826
female_age	.006544	.0018512	3.53	0.000	.0029125	.0101755
_cons	.2982093	.034509	8.64	0.000	.2305132	.3659054

```
6 . test (1.met_online + 1.wave#1.met_online = 0)

 ( 1)  1.met_online + 1.wave#1.met_online = 0

       F(  1,  1374) =   11.50
            Prob > F =    0.0007

7 . reg ed_gap i.wave##i.met_online male_age female_age if missing ==., robust

Linear regression                               Number of obs     =      1,380
                                                F(5, 1374)        =       1.71
                                                Prob > F          =     0.1282
                                                R-squared         =     0.0076
                                                Root MSE          =      2.257
```

ed_gap	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
1.wave	.0031111	.1364243	0.02	0.982	-.2645114	.2707336
1.met_online	-.1852401	.2163052	-0.86	0.392	-.6095643	.2390842
wave#met_online						
1 1	-.2328883	.3024317	-0.77	0.441	-.8261661	.3603894
male_age	.0199352	.0095309	2.09	0.037	.0012385	.0386319
female_age	-.0175527	.010456	-1.68	0.093	-.0380642	.0029587
_cons	-.1790668	.1747749	-1.02	0.306	-.5219213	.1637876

```
8 . test (1.met_online + 1.wave#1.met_online = 0)

 ( 1)  1.met_online + 1.wave#1.met_online = 0

       F(  1,  1374) =    3.77
            Prob > F =    0.0524

9 . reg m_ed_gap i.wave##i.met_online male_age female_age if missing ==., robust

Linear regression                               Number of obs     =      1,380
                                                F(5, 1374)        =       1.46
                                                Prob > F          =     0.1990
                                                R-squared         =     0.0056
                                                Root MSE          =     3.3885
```

m_ed_gap	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
1.wave	.3599106	.2012149	1.79	0.074	-.034811	.7546322
1.met_online	.1620862	.3139907	0.52	0.606	-.4538668	.7780393
wave#met_online						

```
          1 1 |  -.7824095   .4676106    -1.67   0.095    -1.699718    .1348985
              |
     male_age |  -.0044545   .0159131    -0.28   0.780    -.0356711     .026762
   female_age |  -.0056416   .0169297    -0.33   0.739    -.0388524    .0275692
        _cons |   .1915873   .2587269     0.74   0.459    -.3159553    .6991298
------------------------------------------------------------------------------

10 . test (1.met_online + 1.wave#1.met_online = 0)

 ( 1)  1.met_online + 1.wave#1.met_online = 0

       F(  1,  1374) =    3.17
            Prob > F =    0.0753

11 . log close
      name:  <unnamed>
       log:  /Users/lovisaqvarner/Desktop/Main_robust.smcl
  log type:  smcl
 closed on:  30 Apr 2020, 14:38:11
-------------------------------------------------------------------------------
```

Exclusion of unpartnered respondents

```
      name:  <unnamed>
       log:  /Users/lovisaqvarner/Desktop/robust_partner.smcl
  log type:  smcl
 opened on:   9 May 2020, 09:33:52

1 . reg inc_eq i.wave##i.met_online male_age female_age if partnership_status != 3, robust

Linear regression                               Number of obs     =      1,288
                                                F(5, 1282)        =       0.77
                                                Prob > F          =     0.5715
                                                R-squared         =     0.0035
                                                Root MSE          =     .37033

--------------------------------------------------------------------------------
               |               Robust
        inc_eq |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------+----------------------------------------------------------------
        1.wave |   .0340636    .024508     1.39   0.165    -.0140167    .0821439
  1.met_online |   .0120668   .0336352     0.36   0.720    -.0539193    .0780529
               |
wave#met_online |
           1 1 |   -.036434   .0491348    -0.74   0.459    -.1328275    .0599594
               |
      male_age |  -.0014815   .0013545    -1.09   0.274    -.0041387    .0011757
    female_age |   .0003613   .0013776     0.26   0.793    -.0023413     .003064
         _cons |   .1925521   .0319906     6.02   0.000     .1297925    .2553117
--------------------------------------------------------------------------------

2 . test (1.met_online + 1.wave#1.met_online = 0)

 ( 1)  1.met_online + 1.wave#1.met_online = 0

       F(  1,  1282) =    0.47
            Prob > F =    0.4919

3 . reg m_inc_ineq i.wave##i.met_online male_age female_age if partnership_status != 3, robust

Linear regression                               Number of obs     =      1,288
                                                F(5, 1282)        =       5.04
                                                Prob > F          =     0.0001
                                                R-squared         =     0.0203
                                                Root MSE          =     .49039

--------------------------------------------------------------------------------
               |               Robust
    m_inc_ineq |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------+----------------------------------------------------------------
        1.wave |  -.0074472   .0316893    -0.24   0.814    -.0696157    .0547214
  1.met_online |  -.0038099   .0452861    -0.08   0.933    -.0926528     .085033
               |
wave#met_online |
           1 1 |  -.0967617   .0657517    -1.47   0.141    -.2257544     .032231
               |
      male_age |   .0089987   .0020972     4.29   0.000     .0048844     .013113
    female_age |   -.007665   .0022669    -3.38   0.001    -.0121123   -.0032178
         _cons |   .5241664   .0405171    12.94   0.000     .4446793    .6036535
--------------------------------------------------------------------------------

4 . test (1.met_online + 1.wave#1.met_online = 0)

 ( 1)  1.met_online + 1.wave#1.met_online = 0

       F(  1,  1282) =    4.42
            Prob > F =    0.0357

5 . reg f_inc_ineq i.wave##i.met_online male_age female_age if partnership_status != 3, robust

Linear regression                               Number of obs     =      1,288
                                                F(5, 1282)        =       5.01
                                                Prob > F          =     0.0002
```

```
                                                R-squared       =     0.0214
                                                Root MSE        =     .43539
```

f_inc_ineq	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
1.wave	-.0266164	.0275743	-0.97	0.335	-.0807122	.0274793
1.met_online	-.0082569	.0400356	-0.21	0.837	-.0867994	.0702856
wave#met_online						
1 1	.1331958	.0597082	2.23	0.026	.0160592	.2503324
male_age	-.0075172	.0018221	-4.13	0.000	-.0110919	-.0039425
female_age	.0073037	.0019656	3.72	0.000	.0034475	.0111599
_cons	.2832815	.0357899	7.92	0.000	.2130684	.3534947

```
6 . test (1.met_online + 1.wave#1.met_online = 0)

 ( 1)  1.met_online + 1.wave#1.met_online = 0

       F(  1,  1282) =    7.92
            Prob > F =    0.0050

7 . reg ed_gap i.wave##i.met_online male_age female_age if partnership_status != 3, robust

Linear regression                               Number of obs   =      1,299
                                                F(5, 1293)      =       1.46
                                                Prob > F        =     0.1989
                                                R-squared       =     0.0067
                                                Root MSE        =     2.2587
```

ed_gap	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
1.wave	.077309	.1424912	0.54	0.588	-.2022303	.3568483
1.met_online	-.1816433	.213504	-0.85	0.395	-.6004955	.2372089
wave#met_online						
1 1	-.3475603	.3086852	-1.13	0.260	-.9531391	.2580186
male_age	.0131637	.0096386	1.37	0.172	-.0057453	.0320727
female_age	-.0102519	.0105225	-0.97	0.330	-.0308949	.0103911
_cons	-.1844816	.1804631	-1.02	0.307	-.5385142	.1695509

```
8 . test (1.met_online + 1.wave#1.met_online = 0)

 ( 1)  1.met_online + 1.wave#1.met_online = 0

       F(  1,  1293) =    5.46
            Prob > F =    0.0196

9 . reg m_ed_gap i.wave##i.met_online male_age female_age if partnership_status != 3, robust

Linear regression                               Number of obs   =      1,261
                                                F(5, 1255)      =       0.72
                                                Prob > F        =     0.6070
                                                R-squared       =     0.0029
                                                Root MSE        =     3.2952
```

m_ed_gap	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
1.wave	.2372685	.2048341	1.16	0.247	-.1645865	.6391235
1.met_online	.158377	.3143945	0.50	0.615	-.4584197	.7751738
wave#met_online						

```
         1 1  |   -.531596   .4787186    -1.11   0.267    -1.470773    .4075809
              |
    male_age  |  -.0117459   .0152877    -0.77   0.442    -.0417381    .0182462
  female_age  |   .0037605    .016189     0.23   0.816        -.028     .035521
       _cons  |   .1391134   .2674736     0.52   0.603    -.3856313    .6638582
------------------------------------------------------------------------------

10 . test (1.met_online + 1.wave#1.met_online = 0)

 ( 1)  1.met_online + 1.wave#1.met_online = 0

       F(  1,  1255) =    1.04
            Prob > F =    0.3073

11 . log close
      name:  <unnamed>
       log:  /Users/lovisaqvarner/Desktop/robust_partner.smcl
  log type:  smcl
 closed on:   9 May 2020, 09:34:09
-------------------------------------------------------------------------------
```