# The Role of Own-Group Density and Local Social Norms for Ethnic Marital Sorting: Evidence from the UK* 

Dan Anderberg ${ }^{\text {a,b,c }}$, and Alexander Vickery ${ }^{\text {a }}$<br>a) Royal Holloway University of London; b) Institute for Fiscal Studies, London<br>c) CESifo, Munich

April 2, 2021


#### Abstract

We exploit the post-war immigration-induced regional variation in ethnic composition among British-born individuals to study inter-ethnic marriages in the UK. Black and Asian individuals are more likely to marry intra-ethnically in regions where the own ethnicity share is relatively large. In order to disentangle the relative roles played by supply effects, preferences and local social norms we estimate a structural marriage market model that allows for conformity behaviour. Using the estimated model, we make predictions for a set of more recent cohorts whose marital choices are still to be completed.


Keywords: Marriage, Ethnicity, Assortative Matching, Conformity, Social Norms
JEL Classification: D10, J11, J12, J15

[^0]
## I Introduction

As western economies are growing increasingly ethnically diverse through immigration, how minorities integrate is crucial not only to their own experiences of their host countries, but also to public opinion on immigration. Indeed, the very concept of integration remains controversial and a variety of indicators and measures have been proposed (Ager and Strang, 2004). A commonly held view is that marriages between ethnic-minority immigrants and natives is the ultimate proof of integration, stirring the ethnic melting pot and diminishing the significance of group differences (Blau et al., 1984). From a social acceptance perspective, inter-ethnic marriages have been seen as the breaking of the "last taboo" in ethnic and race relations (Qian, 2005). The role of inter-ethnic marriages raises interesting dynamic questions. In particular, as ethnic minorities grow in relative size do inter-ethnic marriages become more or less frequent? What role in this process is played by changing marital opportunities and by evolving social norms respectively?

The sharing of a common culture and identity contributes to the value of marriage, rationalizing observed assortative matching on, for instance, race, religion, and nationality. Indeed, marrying across ethnic lines remains a relatively rare event. In his survey of US marriage trends, Fryer (2007) notes that inter-racial marriages account for approximately 1 percent of White marriages, 5 percent of Black marriages, and 14 percent of Asian marriages. Similarly, the empirical marriage pattern suggests that members of large dominant religious groups Protestants, Catholics and Jews - have strong preferences for having their children identify with their own religious beliefs, thereby creating strong incentives for intra-group marriages (Bisin et al., 2004).

In this paper we study inter-ethnic marriages in the UK among White, Black and Asian British-born individuals, exploiting a strong regional variation in ethnic composition stemming directly from the settlement patterns of the post-war immigrants. ${ }^{1}$ We find that inter-ethnic marriages are more common among Blacks than among Asians. More importantly we find that

[^1]Black and Asian individuals are more likely to be married within their own ethnicity in regions where the density of their own ethnicity is relatively high.

In order to interpret these findings and to identify the separate roles played by population supplies, preferences, and local social norms we estimate a structural marriage market model. The seminal work of Choo and Siow (2006) has provided a workhorse model for empirically implementing Becker's transferable utility model of the marriage market (Becker, 1973). Their framework recasts marriages as choices among discrete "types" and with marital surpluses systematically depending on a couple's type-profile. The framework developed by Choo and Siow (2006) has, over the past decade, been implemented and extended in a variety of directions and applied in a variety of contexts. ${ }^{2}$ One key recent extension by Mourifié and Siow (2017) and Mourifié (2019) has been towards incorporating social preferences. Conceptually there are strong parallels to the literature on discrete choice with social interactions (Brock and Durlauf, 2001), a literature that has been instrumental in empirically operationalizing the notion of social influence and endogenous norms of behaviour (Glaeser and Scheinkman, 2014). Marital choices, in particular in the context of inter-ethnic marriages, may reflect endogenous social norms and conformity behaviour: marrying inter-ethnically may be less of a taboo when others do the same. Hence the current paper will study marital choices within and across ethnic boundaries in the UK using a Choo-Siow style model extended to incorporate endogenous conformity preferences. Such an extension requires a strong source of identification (Galichon and Salanié, 2015) which in our case is provided by the rich regional demographic variation. We adopt a multi-market approach where each region is treated as a separate marriage market and where an individual's type is given by her ethnicity-qualification profile. We specify the form of the joint systematic marriage utility to have two key components (i) a "principal preferences" component - common across all marriage markets - that varies with a couple's type profile, and (ii) an endogenous social norm component that depends the strength of conformity preferences and on how common that particular marriage choice is locally among individuals of the own type.

Our paper makes three contributions. First, we highlight that the ethnic composition of UK-

[^2]born individuals varies substantially across regions, directly reflecting the settlement pattern of the first generation of immigrants arriving to the UK between the mid-1950s and the mid-1970s. We document the marriage patterns within our estimating cohorts of UK-born individuals born between 1965 and 1989, both in the aggregate and across regions.

Second, we estimate a structural marriage model with conformity preferences, identified through the regional variation in population supplies. We find that the principal preferences exhibit significant complementarity in ethnicity as well as in academic qualifications. We also find evidence of strong conformity preferences which in turn implies strong variation in local social norms. We show that the estimated model naturally fits the regional pattern with respect to inter-ethnic marriage frequencies: in areas with relatively larger ethnic minority groups, Whites naturally more frequently marry inter-ethnically, but critically, individuals from the ethnic minorities less frequently marry Whites.

Third, we use the estimated model to predict marriage patterns among individuals born between 1990 and 2006. In these recent cohorts the ethnic Asian minority in particular and the Black minority to a lesser extent are substantially larger than in the estimating cohorts. We find that, as the ethnic minorities grow in terms of their shares of the population, Whites will become more likely to marry inter-ethnically. However, Blacks and Asians will themselves become less likely to marry inter-ethnically. These effects are amplified by the endogenously evolving social norms.

A number of recent papers have adopted structural marriage market modelling in order to study marriages across cultures and borders. Relevant to the UK setting, Marini (2019) draws on Dupuy and Galichon (2014) to allow for multiple continuous traits. One of the traits included in her analysis is a measure of identity, specifically an ethnolinguistic fractionalization (ELF) index based on the country of origin of the mother. The author finds that strong evidence for matching on ELF, which is consistent with strong ethnic matching. However, comparing to the current study, Marini does not consider matching directly on ethnicity as a discrete characteristic and does not account for regional demographic variation or social conformity preferences. Two further recent papers use structural marriage market models to study marriages between natives
and migrants. Focusing on the case of Italy, Adda et al. (2019) use the enlargement of the European Union as a natural experiment to highlight the role played by both cultural distance and legal status. Ahn (2020) studies selection into cross-border marriages between Taiwan and Vietnam, predominantly between Taiwanese men and Vietnamese women.

The rest of the paper is organized as follows. Section II gives a brief overview of post-war immigration into the UK. Section III describes the data that we use and the marriage pattern among the estimating cohorts. Section IV set up the model and outlines how the model is identified. Section V presents the estimation results while section VI highlights our predictions for the more recent cohorts. Section VII concludes.

## II UK Post-War Immigration

Following the conclusion of the second world war the UK had a non-White population of around 30,000 people. By the end of the $20^{t h}$ century this figure was over 3 million. ${ }^{3}$ The post-war rise in immigration can be attributed to a combination of government policy and labour demand. In 1948, the British Nationality Act granted individuals from the British Empire the freedom to live and work in the UK. These extensive rights were in place until the early 1960s when, in response to a perceived heavy influx of immigrants, regulation was significantly tightened. Meanwhile, the post-war Labour government embarked on an extensive nationalization policy, promising full employment, fair wages, and homes for all. Recognizing that reconstruction of the British economy required a large influx of immigrant labour, appeals for new workers were aimed at both Europeans and non-European - mainly from the Caribbean and from the Indian subcontinent. The symbolic starting point for immigration from the Caribbean was the journey of the SS Empire Windrush from Kingston, Jamaica, to Tilbury, Essex, in June 1948, carrying close to 500 West Indians to the UK. This began a large wave of migration now referred to as the "Windrush generation". The majority of immigrants from the Indian subcontinent arrived to Britain during the 1950s and 1960s following Britain's relinquishing in 1947 of its Indian empire in 1947 and the turbulent partition of India into what is known today as the Republic

[^3]of India and the Islamic Republic of Pakistan.
One of the key features of the settlement of ethnic minority immigrants arriving to the UK between the 1950s and the 1970s was its particular geographical pattern. As the post-war immigration was driven primarily by a shortage of labor, both skilled and unskilled immigrants settled in areas where the shortages were perceived to be the largest. Many of the Asian immigrants were attracted to the industrial towns in the East and West Midlands and to London's East end, as well as to the traditional textile producing towns in Yorkshire and the North West. The Caribbean immigrants settled predominantly in London - in areas such as Brixton and Notting Hill - filling labor shortages in London's hospitals and transportation.

Our interest will be in UK-born individuals and we will use the substantial spatial variation in ethnicity across regions, reflecting the settlement pattern of the post-war arrivals.

## III Data

Our aim is to study the marital choices of UK-born individuals primarily in terms of ethnicity. We will further include educational attainment in our analysis to explore if there is a relationship between education and ethnic marital sorting.

For our analysis we want to focus on a set of cohorts who (i) exhibit a substantial and geographically varied presence of UK-born ethnic minorities, and (ii) are old enough to have made their marital choices. To this end, we focus on the birth cohorts 1965-1989. We will use data on individuals, aged 25 or above, from the Quarterly Labour Force Survey (QLFS) - the largest household study in the UK - for the survey years 1996-2015. Our choice of geography is based on the Statistical Regions, specifically Wales, Scotland, and the nine statistical regions of England. ${ }^{4}$

## Sample Population

The QLFS allows us to characterize each individual's ethnicity and educational attainment. Moreover, as the survey interviews all adults in each household, it allows us to measure the

[^4]same characteristics for partners. The ethnic groups that we will consider are Whites ( $W$ ), Blacks ( $B$ ), and Asians $(A)$, based on standard classification by the Office for National Statistics. Educational attainment is broadly classified into two groups: "Low" $(L)$ and "High" $(H)$. We classify an individual as having "low" educational attainment if their highest academic qualification is a GCSE (General Certificate of Secondary Education) or no academic qualification at all. In contrast, we classify an individual as having "high" educational attainment if their highest academic qualification is an A-level (Advanced Level) or higher including a university degree. ${ }^{5}$

For our sample, we will focus on UK-born individuals, born between 1965-1989 and aged 25 and above. The age cutoff ensures that the individuals in the sample have had the necessary time to complete full time education and reach a marriage age. ${ }^{6}$ Pooling the 20 years of the QLFS we then obtain a sample of 203,802 individuals, distributed across ethnicity, qualification and gender as shown in Table 1.

Table 1: Descriptive Statistics of the QLFS Sample

|  | White |  | Black |  | Asian |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Low Qual | High Qual | Low Qual | High Qual | Low Qual | High Qual |
| Males | 44,903 | 49,316 | 447 | 579 | 963 | 1,737 |
| Females | 45,815 | 55,456 | 519 | 845 | 1,179 | 2,043 |

Notes: The sample consists of all UK-born individuals observed in the QLFS 1996-2015, born between 19651989, living in England, Scotland or Wales, and aged 25 or above when observed, and with available information on gender, ethnicity, educational attainment, marital status. A person has a "low qualification" if they either no qualification or a GCSE, and they have a "high qualification" if they have an A-level qualification or higher.

As the QLFS also identifies cohabiting couples, we treat cohabitors as married. Hence being "married" in our context includes both formal marriage and live-in partnerships. In total, 67 percent of the individuals in the sample population are, by this definition, married. It should be noted that we do not impose the same cohort and age-restrictions on partners, though the vast majority of partners do satisfy them. While our key focus will be on marriages between

[^5]UK-born individuals, marriages to non-UK-born spouses are also observed and are known to be particularly common among the low qualified Asians (Charsley et al., 2012). Hence we will include such marriages in our analysis below.

## Geographical Variation

Figure 1 highlights the uneven presence of ethnic Black and Asian across the UK in the estimating cohorts. Panel (a) shows how the ethnic Black are heavily concentrated in the London region, and to a lesser extent in the East and West Midlands. Panel (b) shows how the ethnic Asians are heavily concentrated in London and the West Midlands, and, to a lesser extent, in the North West and in eastern regions. Conversely, this of course implies that proportion Whites (not shown) is relatively low in London, and in the East and West Midlands, and particularly high in regions such as the South West and the North East. As our estimating sample includes only UK-born individuals, this illustrates how the geographical variation of ethnicity among our sample population reflects the settlement patterns of the first generation discussed in Section II. This gives us reassurance that the location choices of the second generation individuals were, in effect, determined exogenously by their parents.

While not shown here, there is also some geographical variation in the distribution of educational attainment with London, the southern regions of England, and England having higher levels of attainment than the rest of England and Wales. ${ }^{7}$

## Empirical Marital Choices

We will define "marital status" to have three categories: single, married to a UK-born partner, and married to a non-UK-born partner. ${ }^{8}$ Figure 2 shows the distribution of marital status by gender, educational attainment and ethnicity for our sample population.

Two key features with respect to ethnicity stand out. First, the rate of singlehood is sub-

[^6]Figure 1: Ethnic Composition by Region


Notes: The sample consists of all UK-born individuals observed in the QLFS 1996-2015, born between 1965-1989, living in England, Scotland or Wales, and aged 25 or above when observed, and with available information on gender, ethnicity, educational attainment, marital status.
stantially higher among Blacks than among Whites and Asians, mirroring findings from the US (Ellwood and Crane, 1990; Saluter, 1994; Brien, 1997; Seitz, 2009). Second, the proportion of individuals who are married to non-UK-born partners is particularly high among Asians. With respect to educational attainment, it is interesting that, for both the Black and the White ethnic groups, being more educated also makes you more likely to be married, whereas the opposite is true for Asians.

Figure 3 shows the distribution of partner type by own "type" - defined by ethnicity and qualification - and gender. ${ }^{9}$ A few things stand out. The vast majority of married White individuals are married to White partners. In contrast, among married Black males, White partners are as common as Black partners. Around 80 percent of married Asians are married to Asian partners. ${ }^{10}$ The figure also highlights that there is educational homogamy: for any ethnicity and gender, a qualified individual is more likely to have qualified spouse than an

[^7]Figure 2: Distribution of Marital Status by Own Type and Gender


Notes: The sample consists of all UK-born individuals observed in the QLFS 1996-2015, born between 1965-1989, living in England, Scotland or Wales, and aged 25 or above when observed, and with available information on gender, ethnicity, educational attainment, marital status. A person has a "low qualification" if they either no qualification or a GCSE, and they have a "high qualification" if they have an A-level qualification or higher.
unqualified individual. Overall, 58 (66) percent of low qualified married males (females) are married to low qualified partners, whilst around 75 (67) percent of high qualified married males (females) have high qualified partners.

More central to our analysis is how the prevalence of intra- versus inter-ethnic marriages varies with the ethnic composition. To give a first indication of this, Figure 4 shows the proportion of intra-ethnic marriages by region, gender and ethnicity. In order to make the figure more interpretable, the regions have been ordered in ascending order in terms of the proportion Whites, starting with London as the most ethnically diverse region through to Wales as the most homogenously ethnically White. Naturally, the proportion of intra-ethnic marriages for Whites is close to 100 percent in regions where the Black and Asian ethnic minorities are very small. The more central feature highlighted by the figure is how the prevalence of intra-ethnic marriages among the ethnic minorities varies systematically across regions. Specifically, the figure suggests that, for both Blacks and Asians, intra-ethnic marriages are more common in areas where each respective ethnic minority is comparatively large. As we will see below, this

Figure 3: Distribution of Partner Type by Own Type and Gender


Notes: The sample consists of all UK-born individuals observed in the QLFS 1996-2015, born between 19651989, living in England, Scotland or Wales, aged 25 or above, married to a UK-born partner at the time of the survey, and with available information on gender, ethnicity, and educational attainment. A person has a "low qualification" if they either no qualification or a GCSE, and they have a "high qualification" if they have an A-level qualification or higher.
will be a central feature of the model that we will estimate and also of the predictions about future marriage behaviour. ${ }^{11}$

In order to assess the robustness of this stylized observation Table 2 presents a set of simple probit regressions. In each reported regression we use the relevant subsample of individuals married to UK-born partners, and regress a dummy for the individual being intra-ethnically married on the proportion of the local marriage market population who are of the own ethnicity. As we have defined an individual's relevant marriage market as comprising all UK-born individuals, born between 1965-1989 and in the own region, we use our full estimating sample to characterize the proportion of marriage market peers who are of the own ethnicity.

The estimated marginal effects from these basic probit regressions are shown in the first

[^8]Figure 4: Proportion Intra-Ethnic Marriages by Region and Gender


Notes: The sample consists of all UK-born individuals observed in the QLFS 1996-2015, born between 1965-1989, living in England, Scotland or Wales, aged 25 or above, married to a UK-born partner at the time of the survey, and with available information on gender, ethnicity, and educational attainment. $m=$ male, $f=$ female.
column for each gender and ethnicity. For Whites the estimated marginal effects of 0.065 for males and 0.074 for females imply that a 15 percentage point reduction in the proportion Whites in the local marriage market (roughly corresponding to the difference between London and Wales) is associated with around a one percentage point reduction in the likelihood of the partner to a married White person being White.

For Blacks the relationship between the proportion intra-ethnically married and the local proportion Blacks is much more dramatic. A ten percentage points increase in the local share of Blacks (again roughly corresponding to the difference between London and a typical area with less than one percent Blacks), is associated with around a 40 percentage points increase in the likelihood of the partner to a married Black person also being Black. Similarly, ten percentage points increase in the share of Asians in the marriage market is associated with around a 20 -

Table 2: The Effect of the Own Ethnicity Share on the Probability of Being Intra-Ethnically Married

|  | (i) | (ii) | (iii) | (i) | (ii) | (iii) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | White Males |  |  | White Females |  |  |
| Pr. Own Eth. | $\begin{gathered} 0.065^{* * *} \\ (0.006) \end{gathered}$ | $\begin{gathered} \hline 0.065^{* * *} \\ (0.006) \end{gathered}$ |  | $\begin{gathered} 0.074^{* * *} \\ (0.007) \end{gathered}$ | $\begin{gathered} \hline 0.074^{* * *} \\ (0.007) \end{gathered}$ |  |
| Pr. Own Born 1940-1960 | $\begin{gathered} 0.752^{* * *} \\ (0.001) \end{gathered}$ |  |  | $\begin{gathered} 0.752^{* * *} \\ (0.001) \end{gathered}$ |  |  |
| Obs. | 61,760 |  |  | 61,844 |  |  |
|  | Black Males |  |  | Black Females |  |  |
| Pr. Own Eth. | $\begin{gathered} 4.235^{* * *} \\ (0.579) \end{gathered}$ | $\begin{gathered} 4.840^{* * *} \\ (0.584) \end{gathered}$ |  | $\begin{gathered} 3.683^{* * *} \\ (0.772) \end{gathered}$ | $\begin{gathered} 4.377^{* * *} \\ (0 . .777) \end{gathered}$ |  |
| Pr. Own Born 1940-1960 | $\begin{gathered} 0.899^{* * *} \\ (0.002) \\ \hline \end{gathered}$ |  |  | $\begin{gathered} 0.896^{* * *} \\ (0.003) \end{gathered}$ |  |  |
| Obs. | 352 |  |  | 237 |  |  |
|  | Asian Males |  |  | Asian Females |  |  |
| Pr. Own. Eth. | $\begin{gathered} 3.234^{* * *} \\ (0.531) \end{gathered}$ | $\begin{gathered} 3.417^{* * *} \\ (0.568) \end{gathered}$ |  | $\begin{gathered} 2.200^{* * *} \\ (0.539) \end{gathered}$ | $\begin{aligned} & 1.960^{* *} \\ & (0.585) \end{aligned}$ |  |
| Pr. 1940-60 | $\begin{gathered} 0.526^{* * *} \\ (0.009) \\ \hline \end{gathered}$ |  |  | $\begin{gathered} 0.522^{* * *} \\ (0.009) \end{gathered}$ |  |  |
| Obs. | 933 |  |  | 955 |  |  |
| Est. Spec. | Probit | IV Probit | 1st Stage | Probit | IV Probit | 1st Stage |

Notes: The sample consists of all UK-born individuals observed in the QLFS 1996-2015, born between 1965-1989, living in England, Scotland or Wales, aged 25 or above and married to a UKborn partner, and with available information on gender, ethnicity, educational attainment, marital status. Specification (i) reports the marginal effects from probit regressions of being married to a spouse of the own ethnicity on the proportion of the local population who are of the own ethnicity ("Prop. Own Ethn."). The latter is measured using the same data but without conditioning on being married. Specification (ii) reports marginal effects from IV probit regressions when the proportion of the local population who are of the own ethnicity is instrumented using the corresponding proportion among all (UK- and non-UK-born) individuals born 1940-1960 as observed in the same QLFS data. Column (iii) reports first stage regressions of the proportion of own ethnicity on the corresponding proportion among those born 1940-1960.

* $p<0.05,{ }^{* *} p<0.01,{ }^{* * *} p<0.001$

30 percentage points increase in the likelihood of the partner to a married Asian person being Asian. These simple regressions thus strongly indicate that, as a stylized fact, the prevalence of intra-ethnic marriages among the ethnic minorities is markedly increasing in the share of the own ethnicity in the local marriage market.

A potential concern is that the current population structure may be affected by selective migration. To check for this we create an instrument that draws on the logic that the settlement pattern of the postwar immigrants was a strong determinant of the local ethnic compositions of our UK-born estimating cohorts. Specifically we measure the share of the own ethnicity among individuals born between 1940 and 1960 and living in the same region. ${ }^{12}$ Instrumenting

[^9]in this way leaves the estimated marginal effects of the own ethnicity share on the likelihood of a marriage being intra-ethnic - reported in second column of each panel - completely unaffected for Whites and largely unaffected for Blacks and Asians. This reflects a very strong first stage in each regression, reported in the third column in each panel. This analysis thus suggests that the empirical relationship between the share of own ethnicity in the local population and the prevalence of intra-ethnic marriages is robust to potential internal migration. Hence, in the structural modelling below, we will take the spatial distribution of ethnic types among the UK-born sample as exogenously given and reflecting the post-war settlement pattern.

## IV The Model

Our model builds on Choo and Siow (2006), whose seminal work showed how the static, frictionless, transferable utility equilibrium model of the marriage market could be recast and estimated as a set of discrete choice problems, connected via equilibrium constraints, that collectively identify the marital surplus structure. Their key innovation was to assume that all individuals belong to some discrete set of observable types and that individuals have preferences over partner type with these preferences having both a deterministic and a random component.

The framework introduced by Choo and Siow (2006) has been subsequently generalized in a variety of directions. Our model is closest in spirit to the generalization by Mourifié and Siow (2017) and Mourifié (2019) who allow for "peer effects". The notion that peers may influence individuals' marital choices is one that has previously been explored in the literature. For instance, Adamopoulou (2012) uses panel data on friendship networks and shows that direct friends influence individuals' partnership formation choices. But peer effects also encompass wider social norms, including what choices are socially approved or, conversely, what behaviours are considered to be taboos. What makes social norms particularly interesting and challenging to study is that they are endogenous equilibrium concepts that can vary across groups and geographical areas. Indeed, one of our innovations is to implement a marriage market model with social preferences in a multi-market setting where social norms vary across regions (Burke

[^10]and Young, 2011).
But before turning to the full empirical model, we will start here by presenting a simple special case that has a full analytical solution. This will provide a set of key insights that will be useful going forward.

## An Illustrative Special Case

The special case considered here imposes three simplifying assumptions relative to the full model below. First, we assume that there are only two types of individuals; hence for now we let the type-space be $X=\left\{x_{1}, x_{2}\right\}$, where we can think of $x_{1}$ as a "majority" ethnic group and $x_{2}$ as a "minority" ethnic group. Second, there are two genders, males $m$ and females $f$, and for now we assume that everyone marries someone of the opposite sex. Third, we impose complete gender symmetry, both in terms of population supplies and in terms of preferences. Let $h(x)$ be the proportion of the population (of either gender) that is of ethnicity-type $x \in X .{ }^{13}$ As $x_{1}$ is the majority type, $h\left(x_{1}\right)>1 / 2$ and $h\left(x_{2}\right)<1 / 2$.

Utility is assumed to be transferable, and we postulate a "principal" joint marital utility function that maps husband-wife type profiles $\left(x^{\prime}, x^{\prime \prime}\right) \in X \times X$ into joint utility $\sigma_{x^{\prime} x^{\prime \prime}}$. We can collect all such $\sigma$-terms in a matrix $\Sigma$, which, in this simple case, has dimension $2 \times$ 2. Furthermore, gender-symmetry implies that $\sigma_{x_{1} x_{2}}=\sigma_{x_{2} x_{1}}$ : in a mixed-ethnicity marriage, the joint utility is the same irrespective of whether it is the husband or the wife who is the majority/minority type. In contrast, there is no restriction on the relationship between $\sigma_{x_{1} x_{1}}$ and $\sigma_{x_{2} x_{2}}$. In addition to the preference parameters in $\Sigma$, there is a conformity-preference parameter $\phi \in[0,1)$ parameterizing the strength of peer effects.

For convenience we will use probability notation to describe the equilibrium. ${ }^{14}$ Hence let $\mu_{x^{\prime \prime} \mid x^{\prime}}^{k}$ denote the equilibrium probability that an individual of gender $k$ and type $x^{\prime}$ marries a partner of type $x^{\prime \prime}$. In the gender-symmetric equilibrium, these probabilities will be genderindependent, but we include the gender superscript for now so as to help characterize the equilib-

[^11]rium. The joint systematic utility when a male of type $x^{\prime}$ marries a female of type $x^{\prime \prime}$, inclusive of peer effects is assumed to take the form,
\[

$$
\begin{equation*}
W_{x^{\prime} x^{\prime \prime}} \equiv \sigma_{x^{\prime} x^{\prime \prime}}+\phi \log \left(\mu_{x^{\prime \prime} \mid x^{\prime}}^{m}\right)+\phi \log \left(\mu_{x^{\prime} \mid x^{\prime \prime}}^{f}\right) . \tag{1}
\end{equation*}
$$

\]

In other words, in addition to the principal utility $\sigma_{x^{\prime} x^{\prime \prime}}$, the joint utility depends on the equilibrium proportions of the own type who make the same choice. ${ }^{15}$ It will be useful to establish some terminology. We will refer to $\phi$ as the "social preferences" (parameter) and the $\phi \log \left(\mu_{x^{\prime} \mid x^{\prime \prime}}^{k}\right)$ terms as the equilibrium "social norms".

In addition, any given individual is assumed to have additional additively separable random utility shocks over possible partner types $x \in X$ which, following the literature, we take to be extreme value distributed and independent across individuals and partner types (see below for details). A well-known feature of the equilibrium, driven by the assumption of additively separable utility shocks over partner type (Chiappori et al., 2017), is that the joint marital utility $W_{x^{\prime} x^{\prime \prime}}$ will be split into a male part, $U_{x^{\prime} x^{\prime \prime}}$, and a female part, $V_{x^{\prime} x^{\prime \prime}}$. The extreme-value distribution of preference shocks in turn implies that the equilibrium choice frequencies will take the standard logit form. Specifically, the relative choice frequencies will satisfy $\log \left(\mu_{x^{\prime \prime} \mid x^{\prime}}^{m} / \mu_{x^{\prime} \mid x^{\prime}}^{m}\right)=U_{x^{\prime} x^{\prime \prime}}-U_{x^{\prime} x^{\prime}}$ for males and $\log \left(\mu_{x^{\prime} \mid x^{\prime \prime}}^{f} / \mu_{x^{\prime \prime} \mid x^{\prime \prime}}^{f}\right)=V_{x^{\prime} x^{\prime \prime}}-V_{x^{\prime \prime} x^{\prime \prime}}$ for females. Adding these together gives that

$$
\begin{equation*}
\log \left(\frac{\mu_{x^{\prime \prime} \mid x x^{\prime}}^{m}}{\mu_{x^{\prime} \mid x^{\prime}}^{m}}\right)+\log \left(\frac{\mu_{x^{\prime} \mid x^{\prime \prime}}^{f}}{\mu_{x^{\prime \prime} \mid x^{\prime \prime}}^{f}}\right)=W_{x^{\prime} x^{\prime \prime}}-\frac{W_{x^{\prime} x^{\prime}}+W_{x^{\prime \prime} x^{\prime \prime}}}{2}, \tag{2}
\end{equation*}
$$

where we used that $U_{x^{\prime} x^{\prime \prime}}+V_{x^{\prime} x^{\prime \prime}}=W_{x^{\prime} x^{\prime \prime}}$ and that, due to gender-symmetry, the utilities from intra-ethnic marriages will be equally shared, $U_{x x}=V_{x x}=W_{x x} / 2$ for either type $x \in X$. We can substitute for the joint systematic utilities in (2) using (1) and simplify using that the

[^12]equilibrium marriage probabilities are gender-independent. ${ }^{16}$ Doing so, and rearranging, yields
\[

$$
\begin{equation*}
\log \left(\frac{\mu_{x^{\prime} \mid x^{\prime}}}{\mu_{x^{\prime \prime} \mid x^{\prime}}} \frac{\mu_{x^{\prime \prime} \mid x^{\prime \prime}}}{\mu_{x^{\prime} \mid x^{\prime \prime}}}\right)=\frac{\Delta}{2}+\phi \log \left(\frac{\mu_{x^{\prime} \mid x}}{\mu_{x^{\prime \prime}} \mid x^{\prime}} \frac{\mu_{x^{\prime \prime} \mid x^{\prime \prime}}}{\mu_{x^{\prime} \mid x^{\prime \prime}}}\right), \tag{3}
\end{equation*}
$$

\]

where $\Delta \equiv \sigma_{x^{\prime} x^{\prime}}+\sigma_{x^{\prime \prime} x^{\prime \prime}}-2 \sigma_{x^{\prime} x^{\prime \prime}}$ and is strictly positive under type-complementarity in the principal joint utilities. Let $r \equiv h\left(x_{2}\right) / h\left(x_{1}\right) \in(0,1 / 2)$ denote the relative frequency of the ethnic minority type. Market balance implies that, in equilibrium, the inter-ethnic marriage frequency for the majority type will be directly proportional to the inter-ethnic marriage frequency for the minority type, $\mu_{x_{2} \mid x_{1}}=r \mu_{x_{1} \mid x_{2}}$. Moreover, since everyone marries someone, $\mu_{x^{\prime} \mid x^{\prime}}=1-\mu_{x^{\prime \prime} \mid x^{\prime}}$, for either type. Hence we can re-write (3) as a single equation characterizing the inter-ethnic marriage rate among the minority type, $\mu_{x_{1} \mid x_{2}}$, as follows,

$$
\begin{equation*}
\frac{r\left(\mu_{x_{1} \mid x_{2}}\right)^{2}}{\left(1-r \mu_{x_{1} \mid x_{2}}\right)\left(1-\mu_{x_{1} \mid x_{2}}\right)}=\exp \left(-\frac{\Delta}{2(1-\phi)}\right) . \tag{4}
\end{equation*}
$$

Equation (4) has a unique solution that characterizes $\mu_{x_{1} \mid x_{2}}$ in terms of $\Delta, \phi$ and $r$, and provides a number of insights, starting with some basic comparative statics. ${ }^{17}$

First, complementarity in the principal preferences, $\Delta>0$, generates positive marital sorting. To see this, note that random matching would imply, $\mu_{x_{1} \mid x_{2}}=h\left(x_{1}\right)$, and would be an equilibrium if and only if $\Delta=0$. In contrast, any $\Delta>0$ implies $\mu_{x_{1} \mid x_{2}}<h\left(x_{1}\right)$. Second, equation (4) highlights how strictly positive conformity preferences $\phi \in(0,1)$ "amplify" the complementarity from the principal preferences: for a given $\Delta>0$, a higher value of $\phi$ reduces $\mu_{x_{1} \mid x_{2}}$. Third, as the left hand side of (4) is increasing in both $r$ and $\mu_{x_{1} \mid x_{2}}$ it follows that $\mu_{x_{1} \mid x_{2}}$ is decreasing in $r$. Hence the model predicts that, as the minority grows as a share of the population, they become less likely to marry inter-ethnically. This latter feature underlies the model's ability to replicate the stylized fact observed above that ethnic minority individuals are less likely to marry inter-ethnically in areas where they make up a larger share of the population.

[^13]Three further points are worth noting before we move to the full empirical model. The first point relates to scaling of the preference parameters. Equation (4) highlights that rescaling $\Delta$ to $\widehat{\Delta}=\lambda \Delta$ and $\phi$ to $\widehat{\phi}=1-\lambda(1-\phi)$ by some arbitrary $\lambda>0$ would leave the equilibrium unaffected. This feature will return in the full model where we show that the preference parameters are only identified up to a scaling factor, implying a need for a normalization.

The second point relates to over-identification restrictions imposed by the model. In particular, note that the relative type frequency $r$ does not feature in equation (3). This means that, even as $r$ varies, the pairwise relative marriage frequencies in the log-term (same on both sides) do not vary. In the full model this property translates into a key specification test.

Third, and finally, the log-terms on two sides of (3) being the same reflects the static nature of the model, in particular the assumption that social norms adjust contemporaneously. But what if norms were more "sticky" or "backward-looking"? Consider for instance the possibility that the current marrying cohort use the marriage behaviour of a previous cohort - when the relative type composition $r$ was different - as reference behaviour. In that case, the choice probabilities on the left hand side of (3) represents the current equilibrium behaviour whereas the corresponding choice probabilities on the right hand side would be reference behaviour of the previous cohort. Perhaps surprisingly, in this simplified case, whether social norms adjust instantaneously or with a "lag" does not matter. This directly reflects the aforementioned testable property: as long as the preferences are stable, the pairwise relative marriage frequencies will also remain stable along any sequence of evolving economics under either norm-formation. ${ }^{18}$ This invariance result with respect to norm formation will not carry over to the generalized environment below as there we will also have singlehood as a choice, and we will also have gender differences in both principal and social preferences (and between marriage and singlehood). Indeed we will

[^14]consider both contemporaneous and backward-looking norms when we simulate the predicted future marriage behaviour in Section VI.

## General Setup

Our estimating model generalizes the above simple illustrative case by having more than two types, by dropping gender symmetry, allowing for singlehood, and by having multiple regional marriage markets. Thus consider an economy consisting of a continuum of men and women. The individuals differ in observable type $x \in X$, where $X$ is a discrete set with $N$ elements. The individuals in the economy are further partitioned into a discrete set of regions, denoted $G$ with typical element $g$. Let $h(x, k, g)$ denote the probability mass function describing the population distribution across types $x \in X$, genders $k=m, f$, and regions $g \in G$.

Each individual faces a choice between marrying and remaining single ("option 0"). Marriage between a male of type $x^{\prime}$ and a female of type $x^{\prime \prime}$ in region $g$ generates a principal systematic utility denoted $\sigma_{x^{\prime} x^{\prime \prime}}^{g}$, and we can collect these terms in region-specific $N \times N$ matrices, $\Sigma^{g}$, for $g \in G$. The principal utility from remaining single is normalized to zero. As noted above, choices also reflect additive individual random preferences over possible partner types $\varepsilon(x)$ and singlehood $\varepsilon(0)$. Following Choo and Siow (2006), these random utilities are all taken to be i.i.d. extreme value distributed across individuals and choice options. ${ }^{19}$

As above, we use a probability notation to describe equilibrium choices. As there are no interactions across regions, an equilibrium occurs region-by-region. Hence let $\mu_{x^{\prime \prime} \mid x^{\prime}}^{g, k}$ be the probability that a person from region $g$ of gender $k$ and type $x^{\prime}$ chooses $x^{\prime \prime} \in X \cup\{0\}$.

In this multi-market environment, we assume that the relevant reference group for a given individual is the set of individuals of the same gender, type and region. Hence we generalize (1), the joint systematic marriage utility, to

$$
\begin{equation*}
W_{x^{\prime} x^{\prime \prime}}^{g} \equiv \sigma_{x^{\prime} x^{\prime \prime}}^{g}+\phi_{1}^{m} \log \left(\mu_{x^{\prime \prime} \mid x^{\prime}}^{g, m}\right)+\phi_{1}^{f} \log \left(\mu_{x^{\prime} \mid x^{\prime \prime}}^{g, f}\right) . \tag{5}
\end{equation*}
$$

[^15]Note that (5) generalizes (1) also by allowing the social preference parameter $\phi$ to differ between men and women and to be specific to marriage. As we now model also singlehood as a choice, we allow for gender-specific social preferences also with respect to this choice, shifting the systematic utility from (the normalized) zero to

$$
\begin{equation*}
U_{x 0}^{g} \equiv \phi_{0}^{m} \log \left(\mu_{0 \mid x}^{g, m}\right), \quad \text { and } \quad V_{0 x}^{g} \equiv \phi_{0}^{f} \log \left(\mu_{0 \mid x}^{g, f}\right), \tag{6}
\end{equation*}
$$

for males and females of type $x$ in region $g$ respectively. In this general setup we thus have four social preference parameters which we can collect in a vector $\Phi \equiv\left\{\phi_{s}^{k}\right\}_{s=0,1}^{k=m, f}$.

As utility is transferable in marriage, it can be shared in any way between the partners. As above, in equilibrium, $W_{x^{\prime} x^{\prime \prime}}^{g}$ will be split into a male part $U_{x^{\prime} x^{\prime \prime}}^{g}$ and a female part $V_{x^{\prime} x^{\prime \prime}}^{g}$. A given male of type $x^{\prime}$ from region $g$ will then make the choice, $x^{\prime \prime} \in X \cup\{0\}$, that maximizes $U_{x^{\prime} x^{\prime \prime}}^{g}+\varepsilon\left(x^{\prime \prime}\right)$ that, in addition to $U_{x^{\prime} x^{\prime \prime}}^{g}$, accounts for his idiosyncratic utility shocks. With the random preferences being i.i.d. extreme value distributed, it follows that the choice frequencies take the standard logit form. Specifically, relative to singlehood, for males, $\log \left(\mu_{x^{\prime \prime} \mid x^{\prime}}^{g, m} / \mu_{0 \mid x^{\prime}}^{g, m}\right)=$ $U_{x^{\prime} x^{\prime \prime}}^{g}-U_{x^{\prime} 0}^{g}$, while for females, $\log \left(\mu_{x^{\prime} \mid x^{\prime \prime}}^{g, f} / \mu_{0 \mid x^{\prime \prime}}^{g, f}\right)=V_{x^{\prime} x^{\prime \prime}}^{g}-V_{0 x^{\prime \prime}}^{g}$.

## Identification

We will here outline how variation in population supplies across regions - which serve as our underlying markets - is central to identification. However, we will start by noting that the model as specified is too general in two key ways: (i) the inclusion of peer effects on all choices means that the preference parameters are only identified up to a scale factor, thus requiring a normalization to be imposed, and (ii) allowing the principal preferences to be unrestricted across regions is too general to be identified.

To see the first part, we can use the logit forms and the fact that $U_{x^{\prime} x^{\prime \prime}}^{g}+V_{x^{\prime} x^{\prime \prime}}^{g}=W_{x^{\prime} x^{\prime \prime}}^{g}$. Substituting using (5) and (6) we obtain that, in equilibrium,
$\sigma_{x^{\prime} x^{\prime \prime}}^{g}=\left(1-\phi_{1}^{m}\right) \log \left(\mu_{x^{\prime \prime} \mid x^{\prime}}^{g, m}\right)+\left(1-\phi_{1}^{f}\right) \log \left(\mu_{x^{\prime} \mid x^{\prime \prime}}^{g, f}\right)-\left(1-\phi_{0}^{m}\right) \log \left(\mu_{0 \mid x^{\prime}}^{g, m}\right)-\left(1-\phi_{0}^{f}\right) \log \left(\mu_{0 \mid x^{\prime \prime}}^{g, f}\right)$.

As we interpret the social preferences as preferences for conformity, we will assume that all four $\phi_{s}^{k}$ are non-negative. It will also be central to equilibrium uniqueness that the conformity preferences are not too strong either. To this end we impose the assumption that each $\phi_{s}^{k}$ falls in the unit interval.

Assumption 1. (Conformity preferences) $\phi_{s}^{k} \in[0,1)$ for each $k=m, f$ and $s=0,1$.
Equation (7) highlights the need for a normalization as it can be multiplied through by an arbitrary constant. Specifically, consider the rescaling by some arbitrary $\lambda>0$, whereby $\widehat{\sigma}_{x^{\prime} x^{\prime \prime}}^{g}=\lambda \sigma_{x^{\prime} x^{\prime \prime}}^{g}$ for each $g \in G$ and type profile $\left(x^{\prime}, x^{\prime \prime}\right) \in X \times X$, and each $\widehat{\phi}_{s}^{k}=1-\lambda\left(1-\phi_{s}^{k}\right)$. Such a rescaling would leave the equilibrium in every region unaffected, reflecting that the preference parameters are only identified up to a scaling factor. A natural approach is to impose a normalization on $\Phi$ and we will return to this below.

The second part reflects the well-known property that the unrestricted Choo and Siow (2006) model with a single market and no peer effects is exactly identified. Specifically, the observed marital choice frequencies across all the regions and types could be perfectly replicated by unrestricted region-specific joint principal marital utility matrices $\left\{\Sigma^{g}\right\}_{g \in G}$ and no peer effects. But that is too general to allow for any meaningful testing and would preclude us from exploring the role of peer effects. Hence it is useful to impose some form of restriction on the principal preferences. The natural first assumption is that they do not vary across regions.

Assumption 2. (Common principal preferences) $\Sigma^{g}=\Sigma$ for all $g \in G$ for some $N \times N$ matrix $\Sigma$.

This assumption is testable. To see this, rearrange (7) and use that, in equilibrium, the market is balanced. This gives us the following form for the matching equations:

$$
\begin{equation*}
\mu_{x^{\prime \prime} \mid x^{\prime}}^{g, m}=\left[\exp \left(\sigma_{x^{\prime} x^{\prime \prime}}^{g}\right)\left(\mu_{0 \mid x^{\prime}}^{g, m}\right)^{1-\phi_{0}^{m}}\left(\mu_{0 \mid x^{\prime \prime}}^{g, f}\right)^{1-\phi_{0}^{f}}\left(\frac{h\left(x^{\prime \prime}, f, g\right)}{h\left(x^{\prime}, m, g\right)}\right)^{1-\phi_{1}^{f}}\right]^{\frac{1}{2-\phi_{1}^{m}-\phi_{1}^{f}}}, \tag{8}
\end{equation*}
$$

for males and

$$
\begin{equation*}
\mu_{x^{\prime} \mid x^{\prime \prime}}^{g, f}=\left[\exp \left(\sigma_{x^{\prime} x^{\prime \prime}}^{g}\right)\left(\mu_{0 \mid x^{\prime}}^{g, m}\right)^{1-\phi_{0}^{m}}\left(\mu_{0 \mid x^{\prime \prime}}^{g, f}\right)^{1-\phi_{0}^{f}}\left(\frac{h\left(x^{\prime}, m, g\right)}{h\left(x^{\prime \prime}, f, g\right)}\right)^{1-\phi_{1}^{m}}\right]^{\frac{1}{2-\phi_{1}^{m}-\phi_{1}^{f}}} \tag{9}
\end{equation*}
$$

for females. Equations (8) and (9) give the following generalization of (3), for the pairwise relative marriage rates,

$$
\begin{equation*}
\log \left(\frac{\mu_{x^{\prime} \mid x x^{\prime}}^{g, k}}{\mu_{x^{\prime \prime} \mid x^{\prime}}^{g, k}} \frac{\mu_{x^{\prime \prime} \mid x^{\prime \prime}}^{g, k}}{\mu_{x^{\prime} \mid x^{\prime \prime}}^{g, k}}\right)=\frac{\sigma_{x^{\prime} x^{\prime}}^{g}+\sigma_{x^{\prime \prime} x^{\prime \prime}}^{g}-\sigma_{x^{\prime} x^{\prime \prime}}^{g}-\sigma_{x^{\prime \prime} x^{\prime}}^{g}}{2-\phi_{1}^{m}-\phi_{1}^{f}} \text { for } x^{\prime}, x^{\prime \prime} \in X \text { and } k=m, f \tag{10}
\end{equation*}
$$

Under the assumption of common principal preferences, the expression on the right hand side of (10) does not vary across regions. Since the probabilities on the left hand side have observable counterparts this implication can be tested and we will do so below.

As our estimation involves solving for the equilibrium at each trial value of the parameter, existence and uniqueness is central to our approach. For this we will draw heavily on Mourifié (2019). Recall also that, within any $g \in G$, adding-up holds: $\mu_{0 \mid x}^{g, k}+\sum_{x^{\prime} \in X}^{g, k} \mu_{x^{\prime} \mid x}^{g, k}=1$ for all $x \in X$ and $k=m, f$. Substituting in these adding-up equations using (8) and (9) generates $2 N$ equations - one for each male and female type - in the $2 N$ unknown singles rates. Mourifié (2019) used this approach to define a mapping from the space of singles rates to itself, which is continuous and where a fixed point corresponds to an equilibrium, allowing the author to prove equilibrium existence using Brouwer's fixed-point theorem. As the current model is a special case, equilibrium existence is guaranteed also in our case. Mourifié (2019) further provides conditions for equilibrium uniqueness. In our special case, uniqueness is guaranteed if $\left(1-\phi_{0}^{k}\right) /\left(2-\phi_{1}^{m}-\phi_{1}^{f}\right)>0$ for $k=m, f$ which is satisfied under Assumption 1.

Having imposed the assumption of common principal preferences, relative marriage frequencies will vary across regions fundamentally due to variation in the relative supply of types. Hence such variation is central to the identification strategy. Furthermore, we need to assume away pathological cases where there is no within-region variation in the supply of types across types and gender.

Assumption 3. (Variation in population supplies) There are at least two regions, $g, g^{\prime} \in G$, such that the population supplies (i) differ across the regions, $h(x, k \mid g) \neq h\left(x, k \mid g^{\prime}\right)$ for some $x \in X$ and $k=m, f$, and (ii) are not constant across types $x \in X$ and genders $k=m, f$ within either region.

Using Assumptions 1-3 we can now turn to identification. To this end, we can collect all preference parameters in a vector, denoted $\boldsymbol{\theta} \equiv\{\Sigma, \Phi\}$. Proposition 1 focuses on the simplest two-by-two case (two types and two regions) and notes that $\boldsymbol{\theta}$ is identified in this case. The proof, which is provided in the online Appendix, shows that in the two-by-two case (where $\boldsymbol{\theta}$ has eight parameters) there are eight matching equations that form a linear equation system of the form $\mathbf{A} \boldsymbol{\theta}=\mathbf{B}$, where $\mathbf{A}$ and $\mathbf{B}$ are a matrix and a vector, respectively, containing only marriage and singles rates (observable in the limit). However, reflecting the need for a normalization, the matrix A has one less than full rank. Consequently, we can identify $\boldsymbol{\theta}$ up to a scale factor (for instance, scaling such that $\widehat{\phi}_{0}^{f}=0$, as we do below.

Proposition 1. (Identification) Suppose $N=2$ and $|G| \geq 2$, and that Assumptions 1 - 3 hold. Then $\boldsymbol{\theta}$ is identified (up to a scale factor) from observable marriage and singles rates.

Proposition 1 shows that two regions is sufficient for identification under the assumption of common principal preferences (Assumption 2). With three or more regions, it becomes possible to relax this assumption. In our empirical application, we will introduce one particularly simple generalization to Assumption 2 that allows us to account in particular for the feature that singles rates are higher in some regions than in others. Specifically, we will introduce region-specific fixed-effects in the marriage utilities such that $\Sigma^{g}=\Sigma+\psi_{g}$ for some $\psi_{g}$ (with $\psi_{g}=0$ for one reference region). This thus introduces a further $|G|-1$ parameters. ${ }^{20}$

As a further generalization that is specific to the current application we include as a further choice option, denoted -1 , marrying a non-UK-born partner. We model this option as having a type- and gender-specific systematic utility - denoted $\sigma_{x,-1}$ for males and $\sigma_{-1, x}$ for females along with its own individual i.i.d. extreme value distributed additive utility component. This thus adds a further $2 N$ parameters. As marrying a partner from outside the economy is - just like singlehood - an observable unilateral choice, the identification of these parameters follows directly from the frequency of this choice relative to the frequency of singlehood.

[^16]
## Empirical Types and Specification Tests

An individual's "type" in our setting is given by their ethnicity and qualification profile. There are three possible ethnicities $Z \equiv\{W, B, A\}$ and two qualification levels $Q \equiv\{L, H\}$. Hence an individual's type is a pair $x=(z, q) \in X \equiv Z \times Q$, and there are $N=6$ types in total. For instance a type $x=(W, L)$ is a White low-qualified individual. Our set of regions $G$ consists of the nine statistical regions of England along with Scotland and Wales, as outlined in Section III. The region-specific population distributions $h(x, k \mid g)$ are taken as given by the observable relative frequencies of ethnicity-qualification-gender types by region.

Before turning to model-estimation we will consider the specification tests implied by (10) under common principal preferences (Assumption 2) with additive region fixed-effects. When $\Sigma^{g}=\Sigma+\psi_{g}$, the right hand side of (10) will be constant across regions. Hence it follows that the left hand side of (10) should also be across regions and, in particular, should not co-vary with any dimension of population supplies. The left hand side of (10) can be viewed as an index of homogamy among individuals of types $x^{\prime}$ and $x^{\prime \prime}$, capturing their relative propensities to marry within- versus across-type and we will use $\log \left(\zeta_{x^{\prime} x^{\prime \prime}}^{g, k}\right)$ to denote it. This homogamy index has a direct empirical counterpart, formed by the corresponding empirical marriage frequencies and henceforth denoted $\log \left(\widetilde{\zeta}_{x^{\prime} x^{\prime \prime}}^{g, k}\right)$. Our specification test thus involves checking that the empirical homogamy index - for each pair of types - is constant across regions in general and unrelated to population supplies in particular.

The proposed specification test has an interesting parallel in the current literature. Recent work by Chiappori et al. (2020), building on Choo and Siow (2006), explores how alternative measures of homogamy proposed in the literature agree with, or not, the notion of stronger preferences for assortative matching as represented by the degree of complementarity in the joint marital utility function. The authors note that many available measures - including popular ones based on random matching as benchmark - fail to separate out the effect of population changes from changes in preferences. The application that Chiappori et al. (2020) focuses on is the much-debated question of whether educational homogamy has increased or decreased over recent decades. However, whereas Chiappori et al. (2020) are interested in measuring changes in
the preferences for assortative matching across cohorts, our interest here lies in testing stability of preferences across regions.

We will approach specification testing in two steps. First we provide here a simple diagnostic version where we plot the empirical homogamy measure across regions and visually check for constancy. However, in doing so we will restrict ourselves to sorting, not on our full types, but on ethnicity and qualifications separately. A benefit of such a visual check is that it helps pinpoint if any particular region stands out. As it turns out, London appears distinct in terms of homogamy. Second, in the online Appendix, we construct a formal test by relating homogamy to population supplies. We will outline this test and summarize the findings below.

As a simple diagnostic test, consider first homogamy among Whites and Blacks. For this ethnic pair, the homogamy index is, following (10), the (log of) the product of (i) the proportion of Whites who are intra-ethnically married relative to the proportion of Whites who are married to Black partners, and (ii) the proportion of Blacks who are intra-ethnically married relative to the proportion of Black individuals who are married to White partners. At population level it would be equivalent to construct the measure using either male- or female-marriage frequencies. ${ }^{21}$ At sample level there will be some gender-variation as some individuals have partners who do not meet the sample cohort and age-restrictions. Hence, for completeness we present the homogamy measure calculated both using male and female empirical marriage frequencies.

Panel (a) in Figure 5 shows the results for Whites and Blacks. The figure shows that the homogamy measure is remarkably stable across regions. The value of the measure is missing for males in Wales and the North-East, and for females in Wales and Scotland. These are cases where we do not observe any intra-ethnic marriages involving Black individuals (see Figure 4). Panel (b) shows the results for Whites and Asians. The homogamy measure here is slightly higher on average, indicating a higher degree of complementarity. Sorting on ethnicity between Whites and Asians appear to have slightly larger regional variation, being somewhat lower in London and the South-West, and possibly higher in the West Midlands. London also stands out clearly when we consider educational homogamy in Panel (c). Here London exhibits substantially

[^17]Figure 5: Ethnic and Educational Homogamy across Regions


Notes: The sample consists of all UK-born individuals observed in the QLFS 1996-2015, born between 1965-1989, living in England, Scotland or Wales, and aged 25 or above when observed, and with available information on gender, ethnicity, educational attainment, marital status. The figure plots $\log \left(\widetilde{\zeta}_{x^{\prime} x^{\prime \prime}}^{g, k}\right)$, the empirical counterpart to the left hand side expression in equation (10), using the observed intra- and inter-marriages between Whites and Blacks (panel a), Whites and Asians (panel b), and low- and high-qualified individuals (panel c), for male and female sample members. The bars indicate 95 percent confidence intervals.
more homogamy than the other regions. In contrast, the educational homogamy is strikingly consistent across the remaining ten regions.

With the exception of London, the diagnostic analysis in Figure 5 provides broad support for the notion of common principal preferences. However, it also highlights the benefit to having a formal test based on the full set of types used in the main estimation. The formal test provided in the online Appendix is based on the implied property that local homogamy should not be related
to any dimension of the population supplies. We test this by regressing $\log \left(\widetilde{\zeta}_{x^{\prime} x^{\prime \prime}}^{g, k}\right)$, for each typepair (and gender), observed across the eleven regions, on each type frequency in $h(\cdot, \cdot \mid g)$. Under the null hypothesis that the model is correctly specified, the estimated regression coefficients should be zero. We form a $\chi^{2}$-test using the sum of the squared standardized coefficients as test statistic. ${ }^{22}$ In line with the simple diagnostics above, we find that the specification test fails when London is included and is not rejected when London is excluded. For this reason, when we explore the robustness of our results, we will present estimates that leave out London.

## Maximum Likelihood Estimator

The model is structurally estimated using maximum likelihood. ${ }^{23}$ The ML-estimation takes the region-specific population distributions as given and solves for the regional equilibria, denoted $\mu_{x^{\prime} \mid x}^{g, k}(\widehat{\boldsymbol{\theta}})$, for every trial value of the parameter vector $\widehat{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is the set of possible (normalized) parameter vectors. In the data we observe $M^{g, k}(x)$ individuals from region $g \in G$, of gender $k=m, f$, and of type $x \in X$, and their empirical choice frequencies $\widetilde{\mu}_{x^{\prime} \mid x}^{g, k}$ for $x^{\prime} \in$ $X \cup\{0,-1\}$. The likelihood contribution of this region, given $\widehat{\boldsymbol{\theta}}$, is

$$
\begin{equation*}
L_{x}^{g, k}(\widehat{\boldsymbol{\theta}})=\prod_{x^{\prime} \in X \cup\{0,-1\}}\left[\mu_{x^{\prime} \mid x}^{g, k}(\widehat{\boldsymbol{\theta}})\right]^{M^{g, k}(x) \widetilde{\mu}_{x^{\prime} \mid x}^{g, k}} \tag{11}
\end{equation*}
$$

and the overall $\log$-likelihood at $\widehat{\boldsymbol{\theta}}$ takes the log of (11) and sums over regions, genders and types. ${ }^{24}$

## V Results

Our estimated model has the following set of parameters: (i) the principal joint utility terms $\sigma_{x^{\prime} x^{\prime \prime}}$ in $\Sigma$, (ii) the social preference parameters, $\left\{\phi_{s}^{k}\right\}_{s=0,1}^{k=m, f}$, (iii) the regional fixed effects,

[^18]$\left\{\psi_{g}\right\}_{g \in G}$, and (iv) the gender and type-specific utilities from marrying a non-UK-born partner, $\left\{\sigma_{x,-1}\right\}_{x \in X}$ for men and $\left\{\sigma_{-1, x}\right\}_{x \in X}$ for women. As outlined above, two normalizations are imposed. First, we normalize one social preference parameter to zero. Specifically, we impose the normalization $\widehat{\phi}_{0}^{f}=0$. Second, we normalize one regional fixed effect to zero. Hence the full version of our estimated model has a total of 61 free parameters.

We can use likelihood ratio tests to test the joint statistical significance of key groups of parameters and we do so in Table A. 1 in the Appendix. There we show that we can reject the restrictions that (i) all the regional fixed effects are zero, and (ii) all the social preference parameters are zero. ${ }^{25}$ These tests thus confirm that there are regional differences in the overall value of marriage relative to singlehood, and that accounting for conformity preferences allows the model to better fit the data.

As the exact scale of the model parameters is arbitrary, we relegate the presentation of their estimated values to the Appendix (Tables A. 2 and A.3). Here we will instead focus on first presenting the model fit to the data, and then what the results imply in terms of complementarities in the joint utility function.

## Model Fit

The model easily fits the aggregate data on marital status and partner type. In the online Appendix, we present model-predicted versions of Figure 2 and Figure 3 which are both very close to the empirical ones. Thus the model naturally replicates the empirical homogamy both on ethnicity and on qualifications at the aggregate level.

More critical for our purposes is how the model fits the local marriage patterns, in particular how marriage choices vary with the local ethnic composition. Figure 6 plots, for each ethnicity and gender, the observed and model-predicted proportions of marriages that are intra-ethnic against the share of own ethnicity in the local population.

The two top panels highlight how, trivially, for White males and females the proportion of intra-ethnic marriages approaches unity as the share of Whites in the local population approaches

[^19]Figure 6: Observed and Predicted Proportion of Intra-Ethnic Marriages Against Own Ethnic Share Across Regions by Ethnicity and Gender


Notes: The sample consists of all UK-born individuals observed in the QLFS 1996-2015, born between 1965-1989, living in England, Scotland or Wales, aged 25 or above, married to a UK-born partner at the time of the survey, and with available information on gender, ethnicity, and educational attainment.
unity. The figures however also show that the model provides a realistic prediction for the two areas (London and West Midlands) where the share of Whites in the local population is substantially lower. Turning to Blacks, the two middle panels highlight how the proportion of Blacks who are married to Blacks increases with the share of Blacks in the local population, and how the model correctly predicts this relationship, including for the exceptional case of London. The bottom two panels do the same for Asians, again highlighting how both in the data and in the model predictions, the proportion of marriages that are intra-ethnic increases with the share of the own ethnicity in the local population.

In section VI below we will use the estimated model to predict marriage patterns for more recently born cohorts who are more ethnically diverse and who are yet to go through the marriage market. To this end, it is useful to show that the model predicts choices well in a marriage market that was not used in the estimation. We can do so by re-estimating the model, leaving out one region, and then comparing the predictions for that region with it included and excluded in the estimation respectively. We present the results from this exercise in the Appendix using the West Midlands region as our left-out region. This is an appropriate choice of region as it is the second most ethnically diverse region (after London) in the estimating cohorts and is, in that respect, more typical of the ethnic composition in the more recent cohorts. As seen from Figure 6, our main estimated model predicts intra-ethnic marriages rates well for the West Midlands region in relation to the overall distribution across regions. Table A. 4 in the Appendix shows that these predictions are only marginally affected by the exclusion of the West Midlands from the estimation. This offer reassurance that the model predictions fair well in an "out-of-sample" marriage market with larger ethnic minorities.

## Complementarities

As the model assumes transferable utility, homogamy is fundamentally driven by complementarities in the joint marital utility function. Hence it is instructive to explore the strength of complementarities with respect to ethnicity and with respect to qualifications. A complication when doing so is that, when looking for instance at complementarities in ethnicity, this can occur for alternative qualification profiles etc.

Hence consider first complementarity with respect to ethnicity. Fix a male-female qualification profile $\left(q_{m}, q_{f}\right) \in Q \times Q$ and a pair of ethnicities $z, z^{\prime} \in Z, z \neq z^{\prime}$, and define

$$
\begin{equation*}
\Delta_{z, z^{\prime} \mid q_{m}, q_{f}} \equiv \sigma_{z, q_{m} ; z, q_{f}}+\sigma_{z^{\prime}, q_{m} ; z^{\prime}, q_{f}}-\sigma_{z, q_{m} ; z^{\prime}, q_{f}}-\sigma_{z^{\prime}, q_{m} ; z, q_{f}}, \tag{12}
\end{equation*}
$$

as the ethnicity complementarity between $z$ and $z^{\prime}$ conditional on husband-wife qualification profile $\left(q_{m}, q_{f}\right)$. As there are four possible qualification profiles and three possible ethnicity pairs, there are a total of twelve conditional ethnicity complementarities.

Similarly, consider complementarity with respect to qualification. Now fix a male-female ethnicity profile $\left(z_{m}, z_{f}\right) \in Z \times Z$ define,

$$
\begin{equation*}
\Delta_{q, q^{\prime} \mid z_{m}, z_{f}} \equiv \sigma_{z_{m}, q ; z_{f}, q}+\sigma_{z_{m}, q^{\prime} ; z_{f}, q^{\prime}}-\sigma_{z_{m}, q ; z_{f}, q^{\prime}}-\sigma_{z_{m}, q^{\prime} ; z_{f}, q}, \tag{13}
\end{equation*}
$$

as the ethnicity complementarity between $q$ and $q^{\prime}$ conditional on $\left(z_{m}, z_{f}\right)$. In this case, $(L, H)$ is the only possible qualification-pair. As there are nine possible ethnicity profiles, there are there a total of nine qualification complementarities.

It should be noted that the complementarities are only identified up to a scale factor: a permissible rescaling of all parameters by some $\lambda>0$ (see section IV above) would similarly change the scale of all the estimated complementarities, but not their signs or relative magnitudes.

Table 3 presents estimates of all ethnicity complementarity measures in Panel A and all qualification complementarity measures in Panel B. All measures in Panel A are positive, indicating complementarity in ethnicity for every qualification profile and ethnicity combination. The estimates for Asian-Black marriages are naturally very imprecise and included mainly for completeness. There are two noteworthy features of the estimates in Panel A. First, the ethnicity complementarities do not show any strong qualification patterns, except for possibly being lower for high-qualified individuals in the case of Whites and Asians. Second, the estimated complementarities are such that, at point estimates, $\widehat{\Delta}_{W, A \mid q_{m}, q_{f}}>\widehat{\Delta}_{W, B \mid q_{m}, q_{f}}$ for every male-female qualification profile $\left(q_{m}, q_{f}\right)$; this suggests that the impetus towards positive ethnic sorting is uniformly stronger between Asians and Whites than among Blacks and Whites.

Table 3: Complementarities in Ethnicity and Qualifications

|  | Panel A: Ethnicity Complementarity |  |  |
| :---: | :---: | :---: | :---: |
|  | Ethnicity Pair |  |  |
| Qual. Profile | White, Black | White, Asian | Black, Asian |
| Low, Low | $5.27^{* * *}$ | $7.38^{* * *}$ | $7.17^{* * *}$ |
|  | $(0.54)$ | $(0.88)$ | $(1.17)$ |
| High, Low | $6.01^{* * *}$ | $8.18^{* * *}$ | 68.1 |
|  | $(0.78)$ | $(1.32)$ | $(>200)$ |
| Low, High | $6.21^{* * *}$ | $8.79^{* * *}$ | 37.3 |
|  | $(0.69)$ | $(1.24)$ | $(>200)$ |
| High, High | $5.63^{* * *}$ | $5.87^{* * *}$ | 38.8 |
|  | $(0.86)$ | $(0.96)$ | $(>200)$ |


|  | Panel B: Qualification Complementarity |  |  |
| :---: | :---: | :---: | :---: |
| Husb. Ethnicity | Wife Ethnicity |  |  |
| White | White | Black | Asian |
|  | $1.31^{* * *}$ | $1.00^{* * *}$ | $2.13^{* * *}$ |
| Black | $(0.07)$ | $(0.12)$ | $(0.11)$ |
|  | $0.31^{* * *}$ | $1.24^{* * *}$ | 31.01 |
| Asian | $(0.05)$ | $(0.17)$ | $(>200)$ |
|  | $1.59^{* * *}$ | 28.5 | $1.46^{* * *}$ |
|  | $(0.09)$ | $(>200)$ | $(0.10)$ |

Notes: Panel A presents the estimates of $\Delta_{z, z^{\prime} \mid q_{m}, q_{f}}$ while Panel B presents the estimates of $\Delta_{q, q^{\prime} \mid z_{m}, z_{f}}$. See text for definitions. Standard errors in parenthesis.
${ }^{*} p<0.05,{ }^{* *} p<0.01,{ }^{* * *} p<0.001$

Similarly, all measures in Panel B are positive indicating complementarity in qualifications for every ethnicity profile. The numbers on the main diagonal - thus involving intra-ethnic marriages - are of similar magnitude. Marriages between Whites and Asians generally exhibit stronger qualification complementarities, whereas the opposite holds for marriages between Whites and Blacks.

## Social Preferences

Our preferred specification provides estimated values for three of the social preference parameters, $\widehat{\phi}_{1}^{m}, \widehat{\phi}_{1}^{f}$ and $\widehat{\phi}_{0}^{m}$ as free parameters (whilst $\widehat{\phi}_{0}^{f}=0$ by normalization). From the LR test provided in Table A. 1 we already know that the social preference parameters are jointly statistically significant. The estimated values are presented in Table A. 3 in the Appendix. The estimated singlehood social preference for males, $\widehat{\phi}_{0}^{m}$, is very close to zero (and thus close to $\widehat{\phi}_{0}^{f}$ ), though somewhat imprecisely estimated. In contrast, the two estimated marriage social pref-
erences, $\widehat{\phi}_{1}^{m}$ and $\widehat{\phi}_{1}^{f}$, are substantially larger and more precisely estimated. This suggests that conformity preferences are stronger in relation to marriage than in relation to singlehood and contribute to shaping how the equilibrium marriage patterns respond to variation in population supplies.

Whilst the social preference parameters are common, social norms will vary across regions. Henceforth we will use $\eta_{x^{\prime} x^{\prime \prime}}^{g}$ to denote the joint marital utility arising from local social norms regarding marriages of husband-wife type profile $\left(x^{\prime}, x^{\prime \prime}\right)$, defined as

$$
\begin{equation*}
\eta_{x^{\prime} x^{\prime \prime}}^{g} \equiv \phi_{1}^{m} \log \left(\mu_{x^{\prime \prime} \mid x^{\prime}}^{g, m}\right)+\phi_{1}^{f} \log \left(\mu_{x^{\prime} \mid x^{\prime \prime}}^{g, f}\right) . \tag{14}
\end{equation*}
$$

Social norms will thus vary across partner type-profiles and across regions.
One might intuitively expect social norms to be more favourable towards inter-ethnic marriages in areas where a given ethnic minority is larger. However, that is not necessarily the case as the two components of (14) tend to move in opposite directions: as the minority group gets larger, members of the majority group will generally more often marry inter-ethnically, but at the same time, the members of the minority group will themselves tend to less frequently marry inter-ethnically.

Recall that the total systematic marriage utility is $W_{x^{\prime} x^{\prime \prime}}^{g}=\sigma_{x^{\prime} x^{\prime \prime}}+\eta_{x^{\prime} x^{\prime \prime}}^{g}+\psi_{g}$. In order to visualize the location and spread of social norms relative to the underlying principal preferences Figure 7 plots each estimated $\widehat{\sigma}_{x^{\prime} x^{\prime \prime}}$ term and the distribution of the corresponding estimated social norms, $\widehat{\eta}_{x^{\prime} x^{\prime \prime}}^{g}$, across regions.

The figure is organized into nine panels, each panel corresponding to a husband-wife ethnicity profile. ${ }^{26}$ Within each panel there are four rows, each corresponding to a qualification profile. The $\widehat{\sigma}_{x^{\prime} x^{\prime \prime}}$ terms, which are replicated from Table A. 2 in the Appendix, are highlighted in blue squares in Figure 7. For marriage profiles where both partners belong to the same ethnic group (the panels on the lead diagonal), there is a distinctive U-shaped pattern in the $\widehat{\sigma}_{x^{\prime} x^{\prime \prime}}$ terms. This reflects the aforementioned qualification-complementarities.

[^20]Figure 7: Principal Marriage Utility and Distribution of Social Norms by Marriage Type-Profile


Notes: The "principal utility" depicts the estimated $\widehat{\sigma}_{x^{\prime} x^{\prime}}$ for each husband-wife type profile while the "social norms" depicts the distribution of $\widehat{\eta}_{x^{\prime} x^{\prime \prime}}^{g}$ across husband-wife type profiles and regions.

More of interest here is the distribution of the social norm terms $\widehat{\eta}_{x^{\prime} x^{\prime \prime}}^{g}$. For each husband-wife type-profile there is a distribution across regions of values of $\widehat{\eta}_{x^{\prime} x^{\prime \prime}}^{g}$. These are illustrated using red dots, with a vertical black line indicating their mean value. A few things are worth noting. First, there is distinct variation in $\widehat{\eta}_{x^{\prime} x^{\prime \prime}}^{g}$ for some key marriage profiles, for instance for WhiteBlack inter-ethnic marriages. Second, while the principal preferences $\widehat{\sigma}_{x^{\prime} x^{\prime \prime}}$ are generally quite similar for White-Asian and White-Black inter-ethnic marriages, social norms are, on average, somewhat stronger against White-Asian marriages than against White-Black marriages.

## Other Preferences

The remaining estimated parameters include the regional fixed effects and the preferences for marrying a non-UK-born spouse. Here we will briefly mention what these estimates show. The estimated values are presented in Table A. 3 in the Appendix.

The region fixed-effects - as they are type-profile-independent additive terms - acts as pref-
erence shift term for marriage versus singlehood and are generally precisely estimated. Of the various regions, London stands out as having a substantially lower marriage utility. This thus allows the model to capture the higher singles rate in the capital.

The preferences for marrying a non-UK-born spouse are also fairly precisely estimated. Moreover, they show the expected pattern of being substantially higher among Asians than among Whites or Blacks, and particularly high among low-qualified Asians.

## Robustness to Sample Selection

Before using the estimated model for prediction purposes, we will mention here two robustness analyses. ${ }^{27}$ First, we explored the robustness of the estimated preferences with respect to lowering the age threshold in the sample selection criteria. In the analysis above, we only included individuals aged 25 or above in order to ensure that nearly everyone would have completed their education. In the robustness check, we lowered this threshold to 20 . Doing so naturally leads to higher observed singles rates. Consequently the estimated marriage utilities were generally smaller (relative to singlehood). Nevertheless, the qualitative features - for instance, complementarity in education and similarity in the principal preferences associated with White-Black and White-Asian intra-ethnic marriages - remain unchanged. Second, we re-estimated the model omitting London. This generally also did not affect the qualitative aspects of the primitive preferences, specifically the complementarity patterns. However, leaving out the most ethnically diverse region from the estimation reduced the precision of the estimated parameters.

## VI Predicted Future Ethnic Homogamy

In this final section we will use our estimated model to predict future ethnic homogamy in the UK. Specifically, we will use the parameters from the model - estimated on cohorts born between 1965 and 1989 - alongside the observed demographic shifts to predict marriage patterns among individuals in a set of "more recent" cohorts, born between 1990 and 2006. In doing so we

[^21]will also explore what role endogenously evolving social norms can be expected to play in this process.

In the online Appendix we present a detailed description of the demographic changes. Most notably, both ethnic minorities have grown as a share of the population with the Asian population growing faster (from just below 3 percent to little over 6 percent) than the Blacks (from around 1.2 percent to about 2.4 percent). The rate of holding a high qualification has also increased in all ethnic groups and both for males and females. ${ }^{28}$

Some of the key mechanisms involved are by now fairly clear. From Figure 6 we know that, as an empirical stylized fact, intra-ethnic minority marriages are more common among Asians and Blacks the larger is the own ethnicity as a share of the local population. In contrast, inter-ethnic marriages were observed to be more common among Whites when the minority shares are larger. The same figure shows that the estimated model naturally replicates these empirical patterns. As a consequence, as the ethnic minority shares have grown in the more recent cohorts, the model will naturally predict that Asians and Blacks will marry intra-ethnically more frequently whilst the majority Whites will inter-marry more frequently. These opposing direct effects will in turn have an ambiguous effect on social norms. With respect to qualifications, our estimated model did not uncover any marked differences in the strength of preferences for ethnic marital sorting across qualification levels (see Table 3). Consequently, the increase in the rate of holding high qualifications between the estimating and recent cohorts can be expected to have at most a minor effect on the future proportion of inter-ethnic marriages.

In Figure 8 we plot predicted shares of intra-ethnic marriages across ethnicity, gender and regions. Indeed, each subfigure presents three sets of predictions. The first set, using square markers, illustrates the model predictions for the estimating cohorts and are thus carried forward from Figure 6 as benchmark. The second set, using diamond-shaped markers, illustrates the predictions for the recent cohorts. Notably, these include predicted equilibrium future social norms.

To highlight the role of the evolving social norms, we also present also a third set of predic-

[^22]tions, using triangular markers, in which each social norm term $\eta_{x^{\prime} x^{\prime \prime}}^{g}$ is fixed at the equilibrium value for the estimating cohorts. The finding that this yields a different set of predictions is a reflection of the fact that, in the full model, it matters whether norms are contemporaneously formed or whether they are backward-looking. This is contrary to the simple illustrative model presented in Section IV which ruled out singlehood and imposed full gender symmetry. In this sense, the difference between the two sets of predictions for the recent cohorts capture, in quantitative terms, the importance of evolving versus "sticky" norms.

Looking first at the Asians (panels e and f), we see that, due to their increasing population shares there is a general movement to the right from the estimating cohorts to the recent cohorts. For instance, whilst in the estimating cohorts Asians made up less than four percent in the nine of the eleven regions, in the recent cohorts, this is the case only in four regions. There is a marked positive direct impact on the predicted rate of intra-ethnic marriages among Asians even holding the social norms fixed. This predicted effect is then further boosted by evolving social norms. As a result, the model predicts that the proportion of intra-ethnic marriages among Asians, for both men and women, in the recent cohorts will be 75 percent or higher in every region.

Turning to the Blacks, Figure 8 (panels c and d) shows that the growth of their population share has been much more modest in all areas outside of London, but nevertheless systematic. Consequently, intra-ethnic marriages among Blacks are also predicted to increase. Moveover, as the these shares increase from relatively low levels, the amplification via evolving social norms becomes relatively large. For London, there has been a sharp increase in the share of Blacks in the local population - from just over 8 percent in the estimating cohorts to over 18 percent in the recent cohorts. But for London, the predicted share of intra-ethnic marriages among Blacks was high already for the estimating cohorts, and is predicted to increase only marginally.

Finally, for the Whites (panels a and b), their population shares were 95 percent or higher in all areas except for London and the West Midlands in the estimating cohorts. In contrast, in the recent cohorts the share of Whites is at 95 percent or higher only in four areas. As a result, their predicted intra-ethnic marriage rates decrease, and this is further amplified by evolving social norms. Nevertheless, the predicted rate of intra-ethnic marriage remains close to or above

Figure 8: Predicted Shares of Intra-Ethnic Marriages Against Own Ethnic Share Across Regions by Ethnicity and Gender in Recent Cohorts Compared to Estimating Cohorts


Notes: The square-shaped markers indicate the model predictions for the "estimating cohorts" born 1965-1989 (carried forward from Figure 6). The diamond-shaped markers indicate the model predictions for the recent cohorts, born 1990-2006, and described in detail in the online Appendix. The triangle-shaped markers indicate the model predictions for the recent cohorts but with social norms held constant at the levels estimated for the estimating cohorts.

95 percent for both men and women in all regions.
Table 4: Predicted Shares of Intra-Ethnic Marriages by Gender, Ethnicity, and Qualification in the Estimating and Recent Cohorts

|  | Panel A: Males |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | White |  | Black |  | Asian |  |
|  | Low Qual. | High Qual. | Low Qual. | High Qual. | Low Qual. | High Qual. |
| Born 1965-1989 | 0.997 | 0.993 | 0.322 | 0.435 | 0.791 | 0.788 |
| Born 1990-2006 | 0.991 | 0.990 | 0.603 | 0.701 | 0.889 | 0.790 |
| Difference | -0.006 | -0.003 | 0.281 | 0.266 | 0.098 | 0.002 |
|  | Panel B: Females |  |  |  |  |  |
|  | White |  | Black |  | Asian |  |
|  | Low Qual. | High Qual. | Low Qual. | High Qual. | Low Qual. | High Qual. |
| Born 1965-1989 | 0.994 | 0.993 | 0.548 | 0.591 | 0.843 | 0.741 |
| Born 1990-2006 | 0.991 | 0.990 | 0.603 | 0.700 | 0.889 | 0.790 |
| Difference | -0.003 | -0.003 | 0.055 | 0.109 | 0.046 | 0.049 |

Notes: Predictions for the estimating cohorts (born 1965-1989) and for recent cohorts (born 1990-2006). A person has a "low qualification" if they either no qualification or a GCSE, and they have a "high qualification" if they have an A-level qualification or higher.

Table 4 aggregates the predictions over regions to show predicted shares of intra-ethnic marriages by gender, ethnicity and qualification level both within the estimating cohorts and the recent cohorts. This highlights how the model predicts that the rate of intra-ethnic marriages will increase in both ethnic minority groups, with a relatively larger increase among Blacks.

Finally, if we aggregate over all marriages in the estimating and the recent cohorts respectively, we find that the predicted fraction of all marriages (between UK-born partners) that are inter-ethnic doubles from about 1.3 percentage points to about 2.5 percentage points. Hence overall as a proportion of all marriages, inter-ethnic marriages are predicted to increase. However, this is still consistent with a predicted decrease in the rates of inter-ethnic marriages among both Blacks and Asians. From Table 4 it is clear that most of this aggregate increase in inter-ethnic marriages comes from changes in marital behaviour for both low- and high-qualified individuals rather than from the increase qualifications between the two cohorts.

## VII Conclusions

As Western economies become more ethnically diverse, will inter-ethnic marriages act as a force for long-term integration, breaking down barriers between natives and immigrants? The UK is a natural example to consider due to its distinct history of post-war Black and Asian immigration. That history created a particular and persistent regional variation in ethnic composition that has carried over to the second and beyond generations of immigrants. Most of the second generation immigrants who were born through the 1960s and 1970s have now gone on to marry and we can study how their marital choices reflect the ethnic composition of their local populations. For example, London is the UK region with by far the largest population shares of both Blacks and Asians. It is then striking that the rate of inter-ethnic marriage is lower among Blacks and Asians in London compared to other regions of the UK; this suggests that higher ethnic minority densities may not foster higher rates of inter-ethnic marriage among members of the minority groups. Furthermore marrying inter-ethnically can also be argued to go against social norms, raising the question of whether, as the ethnic minorities grow in relative size, social norms can be expected to change and contribute to the integration process.

To answer these question we have set up and estimated a structural model of the marriage market, building on the workhorse model of Choo and Siow (2006), and extended with endogenous social norms modelled as conformity preferences. Using that the geographical distribution of UK-born ethnic minorities reflects the settlement patterns of the first generation post-war immigrants, we found that Blacks and Asians are systematically less likely to marry inter-ethnically if they live in regions where their own ethnicity is comparatively larger. The estimates suggest strong ethnic complementarities - stronger than for qualifications - reflected in strong ethnic homogamy. We also found that allowing for endogenous social norms provided a better fit to the data.

We used the estimated model to predict marital patterns among a set of more recent cohorts in order to consider whether more integration through marriage can be expected going forward. Here we found that, although inter-ethnic marriages can be expected to grow as a share of all marriages, members of the ethnic minority groups are predicted to become less likely to marry
inter-ethnically, a result that is also amplified by evolving endogenous social norms. In this sense, the model does not suggest that the Black and Asian populations will become increasingly integrated with the White majority group through the formation of marriages.

A caveat to the current approach is of course the assumption of stable preferences: at the heart of the model is a set of principal preference parameters that describe how the systematic joint marriage utility relate to partner type profile, and these preferences are assumed not to change. Viewed from this perspective, the results presented above can be argued to show that future integration through marriage will require that a change of fundamental preferences. Indeed, the recent work by Merlino et al. (2019), using quasi-random variation in ethnic exposure during childhood, show that exposure to Black peers when young lead whites to have more relationships with Blacks as adults. The authors argue that their results reflect underlying effects on attitudes which, translated to our setting, is more akin to a change in the principal preferences than to a change social norms. It should however be noted that our specification tests detected no systematic differences across the regions in the underlying primitive preferences. This is notable since the members of the estimating cohorts will have had very different exposure to ethnic variation in their youth.

## References

Adamopoulou, E. (2012), 'Peer effects in young adults' marital decisions'. Universidad Carlos III de Madrid.

Adda, J., Pinotti, P. and Tura, G. (2019), ‘There's more to marriage than love: The effect of legal status and cultural distance on intermarriages and separations', Mimeo, Bocconi University.

Ager, A. and Strang, A. (2004), 'Indicators of integration'. Home Office: Development and Practice Report No. 28 .

Ahn, S. Y. (2020), 'Matching across markets: Selection of cross-market matches and its impact on the overall market in the context of cross-border marriage', Mimeo, University of Illinois at Chicago.

Becker, G. (1973), 'A theory of marriage: Part 1', Journal of Political Economy 81(4), 813-846.

Bisin, A., Topa, G. and Verdier, T. (2004), 'Religious intermarriage and socialization in the United States', Journal of Political Economy 112(3), 615-664.

Blau, P., Beeker, C. and Fitzpatrick, K. M. (1984), 'Intersecting social affiliations and intermarriage', Social Forces 62(3), 585-606.

Brandt, L., Siow, A. and Vogel, C. (2016), 'Large shocks and small changes in the marriage market', Journal of the European Economic Association 14(6), 1437-68.

Brien, M. (1997), 'Racial differences in marriage and the role of marriage markets.', Journal of Human Resources 32(4), 741-778.

Brock, W. and Durlauf, S. (2001), 'Discrete choice with social interactions', The Review of Economic Studies 68(2), 235-260.

Burgess, S. (2014), 'Understanding the success of London's schools', CMPO Working Paper No. 14/333.

Burke, M. A. and Young, H. P. (2011), Social norms, in A. Bisin, J. Benhabib and M. Jackson, eds, 'The Handbook of Social Economics', Vol. 1, Elsevier, chapter 8, pp. 311-338.

Charsley, K., Storer-Chuch, B., Benson, M. and Van Hear, N. (2012), 'Marriage-related migration in the UK', International Migration Review 46(4), 861-890.

Chen, L., Choo, E., Galichon, A. and Weber, S. (2019), 'Matching function equilibria: Existence, uniqueness and estimation', Working paper. Available at SSRN, http://dx.doi.org/10.2139/ssrn. 3387335.

Chiappori, P.-A., Costa Dias, M. and Meghir, C. (2020), 'Changes in assortative matching: Theory and evidence for the US', Columbia University.

Chiappori, P.-A., Salanié, B. and Weiss, Y. (2017), 'Partner choice, investment in children, and the marital college premium', American Economic Review 107(8), 2109-21067.

Choo, E. (2015), 'Dynamic marriage matching: An empirical framework', Econometrica 83(4), 1373-1423.

Choo, E. and Siow, A. (2006), 'Who marries whom and why?', Journal of Political Economy 114(1), 175-201.

Dupuy, A. and Galichon, A. (2014), 'Personality traits and the marriage market', Journal of Political Economy 122(6), 1271-1319.

Ellwood, D. and Crane, J. (1990), 'Family change among black Americans: What do we know?', Journal of Economics Perspectives 4(4), 65-84.

Fryer, R. (2007), 'Guess who's been coming to dinner? Trends in interracial marriage over the 20th century', Journal of Economic Perspectives 21(2), 71-90.

Galichon, A. and Salanié, B. (2015), 'Cupid's invisible hand: Social surplus and identification in matching models', Columbia University .

Glaeser, S. and Scheinkman, J. (2014), 'Measuring social interactions', in S. Durlauf and P. Young (eds.) "Social Dynamics", MIT Press .

Graham, B. (2013), 'Comparative static and computational methods for an empirical one-to-one transferable utility matching model', Advances in Econometrics 31, 153-181.

Hansen, R. (2000), Citizenship and Immigration in Post-war Britain: The Institutional Origins of a Multicultural Nation, Oxford University Press.

Marini, A. (2019), 'Who marries whom? The role of identity, cognitive and noncognitive skills in marriage', University of Exeter, Economics Department Discussion Papers Series 19/04.

Merlino, L. P., Steinhardt, M. F. and Wren-Lewis, L. (2019), 'More than just friends? School peers and adult interracial relationships', Journal of Labor Economics . Forthcoming.

Mourifié, I. (2019), 'A marriage matching function with flexible spillover and substitution patterns', Economic Theory 67, 421-461.

Mourifié, I. and Siow, A. (2017), 'The Cobb-Douglas marriage matching function: Marriage matching with peer and scale effects', Working Paper .

Qian, Z. (2005), 'Breaking the last taboo: interracial marriage in America', Contexts 4(4), 33-37.

Saluter, A. (1994), 'Marital status and living arrangements: March 1993.', U.S. Bureau of the Census, Current Population Reports pp. 20-478.

Seitz, S. (2009), 'Accounting for racial differences in marriage and employment.', Journal of Labor Economics 27(3), 385-437.

## Appendix

We first present the results from the likelihood ratio (LR) tests of parameter restrictions. Each column in Table A. 1 corresponds to a specification, and each subsequent column expands on the previous one. Column (i) gives the log likelihood value from a specification without either regional fixed effects or social preferences. Column (ii) reports the log likelihood value when regional fixed effects are included and also gives the details of the LR test based on the improvement in the likelihood from (i) to (ii). The test confirms that regional fixed effects are jointly statistically significant. Column (iii) reports the log likelihood value when we further include the three free social preference parameters. The LR test based on the likelihood-improvement from (ii) to (iii) confirms that these parameters are also jointly statistically significant.

Table A.1: Alternative Empirical Model Specifications

|  | Specification |  |  |
| :--- | :---: | :---: | :---: |
|  | (i) | (ii) | (iii) |
| Log Likelihood Value | $-242,798$ | $-242,036$ | $-241,978$ |
| LR Test Statistic |  | $1,524.1$ | 115.94 |
| Degrees of Freedom |  | 10 | 3 |
| p-value |  | $<0.001$ | $<0.001$ |
| Regional Effects | N | Y | Y |
| Social Preferences | N | N | Y |

Notes: The LR test statistic and associated p-value reported for each specification (ii) - (iii) tests whether the constrained model in the previous column is statistically rejected.

We next present the estimated parameters from our main specification. Table A. 2 presents the estimated principal utility terms $\widehat{\Sigma}$. Table A. 3 presents the remaining estimated preference parameters from the main specification. These include the social preferences $\left\{\widehat{\phi}_{s}^{k}\right\}_{s=0,1}^{k=m, f}$, the regional fixed effects $\left\{\widehat{\psi}_{g}\right\}_{g \in G}$, and the gender-type-specific preferences for marrying a non-UK-born spouse, $\left\{\widehat{\sigma}_{x,-1}\right\}_{x \in X}$ for males and $\left\{\widehat{\sigma}_{-1, x}\right\}_{x \in X}$ for females.

Table A. 4 presents the observed and predicted rates of intra-ethnic marriages, by ethnicity and gender, for the West Midlands region. Two sets of predictions are presented. The first set is based on our full empirical model where the West Midlands region is included in the estimation. The second is based on an estimation where this regions is left out of the estimation.

Table A.2: Estimated Principal Joint Marital Utility by Type Profile

|  |  | Female Type |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | White, Low | White, High | Black, Low | Black, High | Asian, Low | Asian, High |
|  | White, Low | $\begin{gathered} 1.05^{* * *} \\ (0.10) \end{gathered}$ | $\begin{gathered} 1.03^{* * *} \\ (0.10) \end{gathered}$ | $\begin{gathered} \hline-3.42^{* * *} \\ (0.42) \end{gathered}$ | $\begin{gathered} \hline-3.32^{* * *} \\ (0.42) \end{gathered}$ | $\begin{gathered} -2.74^{* * *} \\ (0.43) \end{gathered}$ | $\begin{gathered} -3.12^{* * *} \\ (0.44) \end{gathered}$ |
|  | White, High | $\begin{gathered} 0.61 \\ (0.12) \end{gathered}$ | $\begin{gathered} 1.89^{* * *} \\ (0.12) \end{gathered}$ | $\begin{gathered} \hline-3.47^{* * *} \\ (0.43) \end{gathered}$ | $\begin{gathered} -2.36^{* * *} \\ (0.32) \end{gathered}$ | $\begin{gathered} \hline-3.04^{* * *} \\ (0.48) \end{gathered}$ | $\begin{gathered} -1.29^{* * *} \\ (0.27) \end{gathered}$ |
| Male | Black, Low | $\begin{gathered} -2.55^{* * *} \\ (0.35) \end{gathered}$ | $\begin{gathered} \hline-2.40^{* * *} \\ (0.37) \end{gathered}$ | $\begin{gathered} -1.75^{* * *} \\ (0.26) \end{gathered}$ | $\begin{gathered} \hline-1.72^{* * *} \\ (0.27) \end{gathered}$ | $\begin{gathered} -4.07^{* * *} \\ (0.84) \end{gathered}$ | $\begin{gathered} \hline-3.56 * * * \\ (0.59) \end{gathered}$ |
| Type | Black, High | $\begin{gathered} -2.40^{* * *} \\ (0.34) \end{gathered}$ | $\begin{gathered} -1.94^{* * *} \\ (0.33) \end{gathered}$ | $\begin{gathered} -1.84^{* * *} \\ (0.29) \end{gathered}$ | $\begin{gathered} \hline-0.56^{* * *} \\ (0.17) \end{gathered}$ | $\begin{gathered} -34.4 \\ (>200) \end{gathered}$ | $\begin{gathered} -2.91 * * * \\ (0.49) \end{gathered}$ |
|  | Asian, Low | $\begin{gathered} -2.95^{* * *} \\ (0.41) \end{gathered}$ | $\begin{gathered} -2.61^{* * *} \\ (0.41) \end{gathered}$ | $\begin{gathered} -4.22^{* * *} \\ (0.81) \end{gathered}$ | $\begin{gathered} -33.6 \\ (>200) \end{gathered}$ | $\begin{gathered} 0.64^{* * *} \\ (0.18) \end{gathered}$ | $\begin{gathered} 0.09 \\ (0.19) \end{gathered}$ |
|  | Asian, High | $\begin{gathered} -3.57^{* * *} \\ (0.46) \\ \hline \end{gathered}$ | $\begin{gathered} -1.65 * * * \\ (0.31) \\ \hline \end{gathered}$ | $\begin{gathered} -34.5 \\ (>200) \\ \hline \end{gathered}$ | $\begin{gathered} -35.4 \\ (>200) \end{gathered}$ | $\begin{gathered} 0.13 \\ (0.19) \end{gathered}$ | $\begin{gathered} \hline 1.04^{* *} \\ (0.12) \\ \hline \end{gathered}$ |

Notes: The sample consists of all UK-born individuals observed in the QLFS 1996-2015, born between 1965-1989, living in England, Scotland or Wales, aged 25 or above when observed, and with available information on gender, ethnicity, educational attainment, marital status. A person has a "low qualification" if they have either no qualification or a GCSE, and they have a "high qualification" if they have an A-level qualification or higher. Standard errors in parenthesis.
${ }^{*} p<0.05,{ }^{* *} p<0.01,{ }^{* * *} p<0.001$

Table A.3: Estimates of Remaining Preference Parameters

| Panel A: Social Preferences |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\phi_{1}^{m}$ | $\phi_{1}^{f}$ | $\phi_{0}^{m}$ | $\phi_{0}^{f}$ |  |
|  | $0.55{ }^{* * *}$ | 0.50*** | -0.01 | 0 (ref) |  |
|  | (0.05) | (0.05) | (0.12) | - |  |
| Panel B: Regional-Specific Preferences |  |  |  |  |  |
| $\psi_{\text {London }}$ | $\psi_{\text {West Midl. }}$ | $\psi_{\text {East Midl. }}$ | $\psi_{\text {South East }}$ | $\psi_{\text {North West }}$ | $\psi_{\text {East Engl. }}$ |
| -0.89*** | 0.07 | $0.18{ }^{* * *}$ | -0.10** | -0.11*** | 0.03 |
| (0.07) | (0.04) | (0.04) | (0.03) | (0.04) | (0.04) |
| $\psi_{\text {Yorks. \&Humb }}$ | $\psi_{\text {South West }}$ | $\psi_{\text {North East }}$ | $\psi_{\text {Wales }}$ | $\psi_{\text {Scotland }}$ |  |
| 0.08* | -0.04 | 0 (ref) | 0.01 | -0.19 *** |  |
| (0.04) | (0.04) | - | (0.04) | (0.04) |  |
| Panel C: Preferences for Non-UK-Born Spouse |  |  |  |  |  |
| Males ( $\sigma_{x,-1}$ ) |  |  |  |  |  |
| White, Low | White, High | Black, Low | Black, High | Asian, Low | Asian, High |
| -0.55*** | 0.01 | -0.71*** | -0.14 | 1.03 *** | $0.64 * * *$ |
| (0.11) | (0.10) | (0.13) | (0.09) | (0.15) | (0.11) |
| Females ( $\sigma_{-1, x}$ ) |  |  |  |  |  |
| White, | White, | Black, | Black, | Asian, | Asian, |
| Low | High | Low | High | Low | High |
| $-1.10^{* * *}$ | -0.42** | -1.05*** | -0.68*** | 0.99*** | $0.48{ }^{* * *}$ |
| (0.18) | (0.15) | (0.15) | (0.12) | (0.07) | (0.07) |

Notes: See notes to Table A.1. Standard errors in parenthesis.

Table A.4: Predicted Intra-Ethnic Marriage Rates for the West Midlands Region when Included and Excluded from the Estimation

|  | Males |  |  |  | Females |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | White | Black | Asian |  | White |  | Black | Asian |
| Observed. | 0.992 | 0.571 | 0.926 |  | 0.995 | 0.390 | 0.904 |  |
| Prediction, Incl. | 0.988 | 0.604 | 0.836 |  | 0.991 | 0.419 | 0.853 |  |
| Pred. Error | -0.004 | 0.033 | -0.090 |  | -0.004 |  | 0.029 | -0.051 |
| Prediction, Excl. | 0.988 | 0.609 | 0.818 |  | 0.990 |  | 0.426 | 0.840 |
| Pred. Error | -0.004 | 0.038 | -0.108 |  | -0.005 |  | 0.036 | -0.064 |

Notes: The first row reports the observed rate of intra-ethnic marriages, by gender and ethnicity, in the West Midlands region in the estimating cohorts (born 1965-1989), while the second row shows the corresponding predicted values when using the parameters from the estimation that includes this region. The table then reports the corresponding predictions when using the parameters from an estimation that excludes the West Midlands region.

## Online Appendices

## Appendix B: Proof of Identification

By Assumption 3, we have access to (at least) two regions with non-identical relative population supplies. Without loss of generality, we can focus on exactly two non-identical regions, $g \in$ $G=\left\{g_{1}, g_{2}\right\}$. With two types $X=\left\{x_{1}, x_{2}\right\}$, there are four possible husband-wife type-profiles $\left(x^{\prime}, x^{\prime \prime}\right) \in X \times X$. Hence with two regions, there is a total of $2 N^{2}=8$ matching equations of the form (7). Using the assumption of common principal preferences, $\sigma_{x^{\prime} x^{\prime \prime}}^{g}=\sigma_{x^{\prime} x^{\prime \prime}}$ (Assumption 2) we can conveniently rewrite (7) as follows

$$
\begin{equation*}
\sigma_{x^{\prime} x^{\prime \prime}}+\phi_{1}^{m} \log \mu_{x^{\prime \prime} \mid x^{\prime}}^{g, m}+\phi_{1}^{f} \log \mu_{x^{\prime} \mid x^{\prime \prime}}^{g, f}-\phi_{0}^{m} \log \mu_{0 \mid x^{\prime}}^{g, m}-\phi_{0}^{f} \log \mu_{0 \mid x^{\prime \prime}}^{g, f}=\log \left(\frac{\mu_{x^{\prime \prime} \mid x^{\prime}}^{g, m} \mu_{x^{\prime} \mid x^{\prime \prime}}^{g, f}}{\mu_{0 \mid x^{\prime}}^{g, m} \mu_{0 \mid x^{\prime \prime}}^{g, f}}\right) . \tag{B1}
\end{equation*}
$$

We can collect these matching equations in matrix form by defining, for each region $g \in G$, the following $4 \times 4$ matrix,

$$
A^{g} \equiv\left[\begin{array}{rrrr}
\log \left(\mu_{x_{1} \mid x_{1}}^{g, m}\right) & -\log \left(\mu_{0 \mid x_{1}}^{g, m}\right) & \log \left(\mu_{x_{1} \mid x_{1}}^{g, f}\right) & -\log \left(\mu_{0 \mid x_{1}}^{g, f}\right)  \tag{B2}\\
\log \left(\mu_{x_{2} \mid x_{1}}^{g, m}\right) & -\log \left(\mu_{0 \mid x_{1}}^{g, m}\right) & \log \left(\mu_{x_{1} \mid x_{2}}^{g, f}\right) & -\log \left(\mu_{0 \mid x_{2}}^{g, f}\right) \\
\log \left(\mu_{x_{1} \mid x_{2}}^{g, m}\right) & -\log \left(\mu_{0 \mid x_{2}}^{g, m}\right) & \log \left(\mu_{x_{2} \mid x_{1}}^{g, f}\right) & -\log \left(\mu_{0 \mid x_{1}}^{g, f}\right) \\
\log \left(\mu_{x_{2} \mid x_{2}}^{g, m}\right) & -\log \left(\mu_{0 \mid x_{2}}^{g, m}\right) & \log \left(\mu_{x_{2} \mid x_{2}}^{g, f}\right) & -\log \left(\mu_{0 \mid x_{2}}^{g, f}\right)
\end{array}\right]
$$

and the following $4 \times 1$ column vector,

The $2 N^{2}$ matching equations can be stacked and written succinctly as $\mathbf{A} \boldsymbol{\theta}=\mathbf{B}$ where,

$$
\mathbf{A} \equiv\left[\begin{array}{cc}
I_{4} & A^{1}  \tag{B4}\\
I_{4} & A^{2}
\end{array}\right], \quad \text { and } \quad \mathbf{B} \equiv\left[\begin{array}{c}
B^{1} \\
B^{2}
\end{array}\right]
$$

and where $I_{n}$ is the $n \times n$ identity matrix, and where the parameter vector $\boldsymbol{\theta}$ in this case is

$$
\begin{equation*}
\boldsymbol{\theta} \equiv\left[\sigma_{x_{1} x_{1}}, \sigma_{x_{1} x_{2}}, \sigma_{x_{2} x_{1}}, \sigma_{x_{2} x_{2}}, \phi_{1}^{m}, \phi_{0}^{m}, \phi_{1}^{f}, \phi_{0}^{f}\right]^{\prime} . \tag{B5}
\end{equation*}
$$

Note that the dimensions of $\mathbf{A}$ is $8 \times 8$ and $\mathbf{B}$ is $8 \times 1$ and that $\mathbf{A}$ and $\mathbf{B}$ only contain (log-values of) population marriage and singles rates. These rates all have empirical counterparts, which by standard arguments, converge in probability to their population values as the sample size grows. Hence, $\mathbf{A}$ and $\mathbf{B}$, can be taken to be known.

Remark 1. It is instructive to note how the current setting generalizes the original Choo and Siow (2006) setting with just one region and no peer-effects. In that case the reduced parameter vector would be $\boldsymbol{\theta}_{C S} \equiv\left[\sigma_{x_{1} x_{1}}, \sigma_{x_{1} x_{2}}, \sigma_{x_{2} x_{1}}, \sigma_{x_{2} x_{2}}\right]^{\prime}$ and the matching equations written in matrix form would be $\mathbf{I}_{2} \boldsymbol{\theta}_{C S}=\mathbf{B}$, which trivially just notes that $\boldsymbol{\theta}_{C S}=\mathbf{B}$.

This then suggests that $\boldsymbol{\theta}$ could be obtained by inverting A. However, A does not have full rank. This is consistent with $\boldsymbol{\theta}$ being identified only up to a scaling factor: as it not possible to identify the scale of $\boldsymbol{\theta}$, one parameter can be arbitrarily fixed (e.g. $\phi_{0}^{f}=0$ ). We will correspondingly show that, generically, $\operatorname{rank}(\mathbf{A})=7$. That is, we will show that there is one linear dependence among the eight matching equations.

To do this, we can use basic Gauss-Jordan elimination. As a first step, we can subtract the upper half of $\mathbf{A}$ from the lower half to obtain

$$
\mathbf{A}^{\prime} \equiv\left[\begin{array}{cc}
I_{4} & A^{1}  \tag{B6}\\
0_{4 \times 4} & \Delta_{A}
\end{array}\right]
$$

where $\Delta_{A} \equiv A^{2}-A^{1}$, contains all log-differences in marriages and singles rates across the regions
and where $0_{n \times m}$ is the $n \times m$ matrix with only zeros. Fully written out,

Remark 2. Generically, $\operatorname{rank}\left(A^{g}\right)=4$ for each $g \in G$. A case where this would fail to hold is the pathological case when the supply of types in region $g$ is constant across types and gender. Hence this case is ruled out by the second part of Assumption 3.

If the two regions $g \in\left\{g_{1}, g_{2}\right\}$ had not only common principal preferences (Assumption 2), but also identical relatively population supplies (violating Assumption 3), the equilibrium marriage and singles-rates would be identical in the two regions for each gender, type, and type-profile, whereby $A^{1}=A^{2}$ and $B^{1}=B^{2}$. It then trivially follows that $\mathbf{A}$ would only have $\operatorname{rank}(\mathbf{A})=4$ and the model with peer effects would not be identified.

As $\mathbf{A}^{\prime}$ was obtained from $\mathbf{A}$ via row subtraction, $\operatorname{rank}(\mathbf{A})=\operatorname{rank}\left(\mathbf{A}^{\prime}\right)$ and we will prove that the lower half of $\mathbf{A}^{\prime}$ does not have linearly independent rows. Specifically, we will demonstrate that $\operatorname{rank}\left(\Delta_{A}\right)=3$. For notational simplicity, let $\delta_{i j}$ denote the $(i, j)$-th element of $\Delta_{A}$. Two rounds of Gauss-Jordan elimination, pivoting on rows 1 and 2 of $\Delta_{A}$ in turn yields,

To prove that $\operatorname{rank}\left(\Delta_{A}\right)=3$ we will show that the last two rows of $\Delta_{A}^{\prime}$ are identically the same. We start with the two final terms in column 3 of $\Delta_{A}^{\prime}$. A simple rearranging shows that
these terms are identical if and only if

$$
\begin{equation*}
\frac{\left(\delta_{42}-\delta_{41} \frac{\delta_{12}}{\delta_{11}}\right)-\left(\delta_{32}-\delta_{31} \frac{\delta_{12}}{\delta_{11}}\right)}{\delta_{22}-\delta_{21} \frac{\delta_{12}}{\delta_{11}}}=\frac{\left(\delta_{43}-\delta_{41} \frac{\delta_{13}}{\delta_{11}}\right)-\left(\delta_{33}-\delta_{31} \frac{\delta_{13}}{\delta_{11}}\right)}{\delta_{23}-\delta_{21} \frac{\delta_{13}}{\delta_{11}}}, \tag{B9}
\end{equation*}
$$

holds. We will show that (B9) indeed holds as both sides are equal to unity.
Consider first the left hand side of (B9) and note that this is unity if and only if

$$
\begin{equation*}
\frac{\delta_{31}+\delta_{21}-\delta_{41}}{\delta_{11}}=\frac{\delta_{32}+\delta_{22}-\delta_{42}}{\delta_{12}} . \tag{B10}
\end{equation*}
$$

But (B10) holds as both sides are equal to unity: using the definition of terms in $\Delta_{A}$ in (B7), equation (10) and Assumption 2 immediately imply that $\delta_{11}-\delta_{21}=\delta_{31}-\delta_{41}$ and, directly through their definitions, $\delta_{12}=\delta_{22}$ and $\delta_{32}=\delta_{42}$.

Consider then the right hand side of (B9) and note that this is unity if and only if

$$
\begin{equation*}
\frac{\delta_{33}+\delta_{23}-\delta_{43}}{\delta_{13}}=\frac{\delta_{31}+\delta_{21}-\delta_{41}}{\delta_{11}} . \tag{B11}
\end{equation*}
$$

But (B11) also holds as both sides are unity: as already confirmed above, $\delta_{11}-\delta_{21}=\delta_{31}-\delta_{41}$, while equation (10) and Assumption 2 imply that $\delta_{13}-\delta_{23}=\delta_{33}-\delta_{43}$. A corresponding set of steps show that the final two terms in column 4 of $\Delta_{A}^{\prime}$ are identical.

Equality of the final two rows of $\Delta_{A}^{\prime}$ implies that $\operatorname{rank}\left(\Delta_{A}^{\prime}\right)=3$. Generically, with the two regions having different relative population supplies, there is no further linear dependence between the rows of $A^{1}$ and $A^{2}-A^{1}$, and hence $\operatorname{rank}(\mathbf{A})=\operatorname{rank}\left(\mathbf{A}^{\prime}\right)=7$.

Remark 3. The demonstrated equality of the two final rows of $\Delta_{A}^{\prime}$ implies that the row echelon form of $\mathbf{A}$ is

$$
\operatorname{ref}(\mathbf{A})=\left[\begin{array}{cc}
I_{7} & \widehat{A}_{7 \times 1}  \tag{B12}\\
0_{1 \times 7} & 0_{1 \times 1}
\end{array}\right],
$$

where $\widehat{A}$ is a non-zero vector with indicated size.
The one-less-than-full rank of $\mathbf{A}$ reflects that the equilibrium is invariant to a rescaling of all parameters by a scaling factor $\lambda>$ as outlined in the text. Consequently, one could impose
$\widehat{\phi}_{0}^{f}=0$ as normalization, drop one equation, and solve the remaining seven equations for the remaining seven parameters.

Remark 4. Extending to $N$ types expands the dimensions of $A^{g}$ to $N^{2} \times 4$ and hence expands the dimension of A to $2 N^{2} \times\left(N^{2}+4\right)$. Still, as the parameters can be identified only up to a scaling factor, $\operatorname{rank}(\mathbf{A})=N^{2}+3$. In particular, in this generalization, the echelon form of $\mathbf{A}$ is

$$
\operatorname{ref}(\mathbf{A})=\left[\begin{array}{cc}
I_{N^{2}+3} & \widehat{A}_{\left(N^{2}+3\right) \times 1}  \tag{B13}\\
0_{\left(N^{2}-3\right) \times\left(N^{2}+3\right)} & 0_{\left(N^{2}-3\right) \times 1}
\end{array}\right],
$$

allowing for the identification of $\Sigma$ and, after normalization, three peer effects parameters. At the same time, expanding to more types leads to more testable restrictions based on (10).

Remark 5. Extending to $|G| \geq 3$ regions and allowing for region fixed-effects such that $\Sigma^{g}=$ $\Sigma+\psi_{g}$ further modifies the matching equation system. With (for simplicity) $N=2$ types and $|G|$ regions, the extended parameter vector can be written as

$$
\begin{equation*}
\boldsymbol{\theta}_{F E} \equiv\left[\sigma_{x_{1} x_{1}}, \sigma_{x_{1} x_{2}}, \sigma_{x_{2} x_{1}}, \sigma_{x_{2} x_{2}}, \phi_{1}^{m}, \phi_{0}^{m}, \phi_{1}^{f}, \phi_{0}^{f}, \psi_{1}, \ldots, \psi_{|G|}\right]^{\prime} \tag{B14}
\end{equation*}
$$

and the matching equations can be written in matrix form as $\mathbf{A} \boldsymbol{\theta}_{F E}=\mathbf{B}$ where

$$
\mathbf{A} \equiv\left[\begin{array}{ccc}
I_{4} & A^{1} & C_{1}  \tag{B15}\\
\vdots & \vdots & \vdots \\
I_{4} & A^{|G|} & C_{|G|}
\end{array}\right], \quad \text { and } \quad \mathbf{B} \equiv\left[\begin{array}{c}
B^{1} \\
\vdots \\
B^{|G|}
\end{array}\right]
$$

where $C_{g}$ is the $4 \times|G|$ matrix which has ones in column $g$ and zeros everywhere else. In this case, the dimensions of $\mathbf{A}$ is $4|G| \times(8+|G|)$. But the rank of $A$ is $\operatorname{rank}(A)=N^{2}+2+|G|$. For instance, with $|G|=3$ regions, we have 12 matching equations, nine of which are linearly independent, and we correspondingly have three testable restrictions of the form (10).

## Appendix C: Specification Tests Based on Full Type

In this Appendix we will present results from specification tests based on full type $x \in X$. Recall that under the assumption of "common principal preferences" (Assumption 2), and its extension to regional fixed effects, the right hand side of (10) is independent of $g$. Specifically, the degree of homogamy between any two types, $x^{\prime}$ and $x^{\prime \prime}$, in any region $g$ should reflect the underlying common preferences for assortative mating, not the local type distribution.

To test this model implication, we define the empirical counterpart to the left hand side of (10) as our observed measure of homogamy,

$$
\begin{equation*}
\log \left(\widetilde{\zeta}_{x^{\prime} x^{\prime \prime}}^{g, k}\right)=\log \left(\frac{\widetilde{\mu}_{x^{\prime} \mid x}^{g, k}}{\widetilde{\mu}_{x^{\prime \prime} \mid x^{\prime}}^{g, k}} \frac{\widetilde{\mu}_{x^{\prime}}^{g, k}, x^{\prime \prime}}{\widetilde{\mu}_{x^{\prime} \mid x^{\prime \prime}}^{g, k}}\right) . \tag{C1}
\end{equation*}
$$

We construct this homogamy measure for each pair of types $\left(x^{\prime}, x^{\prime \prime}\right) \in X \times X$, and for each region $g$ and gender $k$. For a given type-profile ( $x^{\prime}, x^{\prime \prime}$ ) and gender $k$ we thus have $|G|$ observations on $\log \left(\widetilde{\zeta}_{x^{\prime} x^{\prime \prime}}^{g, k}\right)$, which, if the model is correctly specified, should not be related to any dimension of $h(\cdot, \cdot \mid g)$. This can be tested by running simple regressions of $\log \left(\widetilde{\zeta}_{x^{\prime} x^{\prime \prime}}^{g, k}\right)$ on each element of $h(\cdot, \cdot \mid g)$. Under the null hypothesis that the model is correctly specified, the $\beta$-coefficients from these regressions should be zero. Note that with six types in $X$, there are fifteen unique typepairs. With two genders and twelve dimensions $h(\cdot, \cdot \mid g)$ (six types for two genders), we can, in principle, estimate principle $360 \beta$-coefficients measuring correlations between local homogamy and local population supplies.

In practice we do not consider homogamy between Black and Asian types as such marriages are too rare in the data for any meaningful analysis. This reduces the number of type pairs that we consider from fifteen to eleven. Furthermore, as the type frequencies sum to unity, there are only eleven independent dimensions in $h(\cdot, \cdot \mid g)$. The standard errors on the estimated $\beta$ coefficients should further account for sampling variation in the empirical homogamy measures. To this end, we compute bootstrapped standard errors that resamples the data to capture the sampling variation in $\log \left(\widetilde{\zeta}_{x^{\prime} x^{\prime \prime}}^{g, k}\right)$.

In order to test the null hypothesis that the locally observed homogamy is unrelated to the
local type distributions, we perform a $\chi^{2}$-test where our test statistic is the sum of the squared (standardized) $\beta$-coefficients. This uses that, under the null hypothesis, the $\beta$-coefficients are approximately normally distributed with zero mean and independent. Hence our most general specification test is a $\chi^{2}$-test with $242(=11 \times 11 \times 2)$ degrees of freedom, testing for the correlation of homogamy across eleven type profiles, two genders and the population supply of eleven gender-types.

Table C.1: Specification Tests

|  | (i) | (ii) | (iii) |
| :--- | :---: | :---: | :---: |
| $\chi^{2}$-test statistic | $298^{* *}$ | 253.7 | 102.2 |
| d.f. $(n)$ | 242 | 242 | 88 |
| p-value | 0.023 | 0.289 | 0.143 |
| London incl. | Yes | No | No |

Notes: The reported $\chi^{2}$ test statistic is the sum of squared standardized $\beta$-coefficients from $n$ simple regressions of local homogamy $\log \left(\widetilde{\zeta}_{x^{\prime} x^{\prime \prime}}^{g, k}\right)$ on local type frequencies $h\left(x^{\prime \prime \prime}, k^{\prime} \mid g\right)$. All three specifications include all non-Black-Asian type pairs $\left(x^{\prime}, x^{\prime \prime}\right)$. Specification (i) includes all type dimensions in $h\left(x^{\prime \prime \prime}, k^{\prime} \mid g\right)$ except for the left-out type of high qualified Asian females. Specification (iii) includes only type dimensions $x^{\prime \prime \prime} \in\left\{x^{\prime}, x^{\prime \prime}\right\}$.

The result from this general test is highlighted in column (i) of Table C.1. This test rejects the null hypothesis of common preferences. From Figure 5 we know that homogamy in London stands out as being potentially different from that in the other regions, not least with respect to qualifications. Moreover, in terms of population supplies, the qualification rate as well as the ethnic minority shares are larger in London than in the other regions. Column (ii) shows that when we omit London from our set of regions, the null hypothesis is indeed no longer rejected.

The general version our test includes the correlation of local homogamy between any two types $\left(x^{\prime}, x^{\prime \prime}\right)$ (excluding matches between Black and Asian types) and the supply of any type $x^{\prime \prime \prime} \in X$. A narrower test, presented in column (iii) focuses specifically on whether the sorting among types $x^{\prime}$ and $x^{\prime \prime}$ depends on the supply frequency of these types only. In this case too the test fails to reject the null hypothesis.

## Appendix D: Estimation

The estimated parameters include $\Sigma$ and $\Phi$. As noted in Section IV we further include as a choice option - denoted -1 - that an individual marries a spouse who is not UK-born and we allow for gender-type-specific preferences for this option. This adds $2 N$ additional parameters. Finally, we allow for region-fixed-effects which adds one further parameter, $\psi_{g}$, for each group (except for the reference region).

The available data is taken to be a random sample from the population. The data on individuals from region $g \in G$ is used to characterize the region-specific type-distributions $h(x, k \mid g)$ which are taken as given throughout the estimation. The data further characterizes the empirical choice-frequencies $\widetilde{\mu}_{x^{\prime} \mid x}^{g, k}$ for all $g \in G$, both genders $k=m, f$, all own-types $x \in X$, and all choice-options, $x^{\prime} \in X \cup\{0,-1\}$. Implementing the ML-estimator requires computing the corresponding model-predicted equilibrium choice-frequencies at any candidate parameter vector $\widehat{\boldsymbol{\theta}}$, denoted $\mu_{x^{\prime} \mid x}^{g, k}(\widehat{\boldsymbol{\theta}})$. Hence we next outline the procedure used for solving for these equilibrium choice-frequencies given $\widehat{\boldsymbol{\theta}}$.

## Computing the Model-Predicted Choice Frequencies

Consider a trial value $\widehat{\boldsymbol{\theta}}$ of the parameter vector. In order to compute the likelihood value associated with $\widehat{\boldsymbol{\theta}}$ we solve for the equilibrium choice frequencies of all types in all regions. Since regions do not interact, the equilibria are solved region-by-region. Below we describe the simple Newton algorithm used.

## Within-Region Algorithm

Fix a group $g \in G$.
Step 0. Guess a single rate for each gender and type, $\mu_{0 \mid x}^{g, k}$, and place in a $2 N$-vector $\mu_{0}^{g}=\left[\mu_{0 \mid x_{1}}^{g,}, \ldots, \mu_{0 \mid x_{N}}^{g, m}, \mu_{0 \mid x_{1}}^{g, f}, \ldots, \mu_{0 \mid x_{N}}^{g, f}\right]^{\prime}$.

Step 1. Given $\widehat{\boldsymbol{\theta}}$, and given $\mu_{0}^{g}$, compute the candidate gender-specific marriage rates of all types to all types using (8) and (9) and also the rates of marriages to non-UK-born partners
using

$$
\begin{equation*}
\mu_{-1 \mid x_{i}}^{g, m}=\left[\exp \sigma_{x_{i},-1}\left(\mu_{0 \mid x_{i}}^{g, m}\right)^{1-\phi_{0}^{m}}\right]^{\frac{1}{1-\phi_{1}^{m}}}, \text { and } \mu_{-1 \mid x_{j}}^{g, f}=\left[\exp \sigma_{-1, x_{j}}\left(\mu_{0 \mid x_{j}}^{g, f}\right)^{1-\phi_{0}^{f}}\right]^{\frac{1}{1-\phi_{1}^{f}}}, \tag{D1}
\end{equation*}
$$

for males and females respectively.
Step 2. Compute the implied gender-type-specific "excess demands" as the deviation from the adding-up conditions,

$$
\begin{equation*}
\Delta_{x_{i}}^{g, m}=\sum_{x_{j} \in X \cup\{0,-1\}} \mu_{x_{j} \mid x_{i}}^{g, m}-1, \text { and } \Delta_{x_{j}}^{g, f}=\sum_{x_{i} \in X \cup\{0,-1\}} \mu_{x_{i} \mid x_{j}}^{g, f}-1 . \tag{D2}
\end{equation*}
$$

Stack the excess demands in a $2 N$-vector $\Delta^{g}=\left[\Delta_{x_{1}}^{g, m}, \ldots, \Delta_{x_{N}}^{g, m}, \Delta_{x_{1}}^{g, f}, \ldots, \Delta_{x_{N}}^{g, f}\right]^{\prime}$. The equilibrium is characterized by zero excess demand for all types and both genders, $\Delta^{g}=0$. Let $\|\cdot\|_{\infty}$ denote the uniform norm. If the candidate marriage/singles rates involve no excess demand, i.e. $\left\|\Delta^{g}\right\|_{\infty} \leq \epsilon$ for a sufficiently small $\epsilon$ (in our case $\epsilon=10^{-5}$ ), terminate the algorithm and take the candidate marriage/singles rates to be the equilibrium rates in region $g$. Otherwise, go to step 3.

Step 3. Update the guess for the singles rates, $\mu_{0}^{g}$, using a simplified Newton step. Let $J$ denote the diagonal $2 N \times 2 N$ matrix where the diagonal terms are the partial derivatives of each excess demand terms $\Delta_{x}^{g, k}$ with respect to the "own" singles rate $\mu_{0 \mid x}^{g, k}$, that is

$$
\begin{equation*}
\frac{\partial \Delta_{x}^{g, k}}{\partial \mu_{0 \mid x}^{g, k}}=1+\sum_{x^{\prime} \in X \cup\{-1\}} \frac{\partial \mu_{x^{\prime} \mid x}^{g, k}}{\partial \mu_{0 \mid x}^{g, k}}, \tag{D3}
\end{equation*}
$$

with $\partial \mu_{x^{\prime} \mid x}^{g, k} / \partial \mu_{0 \mid x}^{g, k}$ and $\partial \mu_{-1 \mid x}^{g, k} / \partial \mu_{0 \mid x}^{g, k}$ obtained from differentiating (8) and (9) and (D1) respectively. Hence $J$ is the Jacobian of $\Delta^{g}$ with respect to $\mu_{0}^{g}$ but ignoring cross-partials. The singles rates are then updated using the Newton step

$$
\begin{equation*}
\mu_{0}^{g^{\prime}}=\mu_{0}^{g}+\kappa J^{-1} \Delta^{g} \tag{D4}
\end{equation*}
$$

where $\kappa \in(0,1)$ is a "dampening" factor. Go back to step 1 with the updated guess, and iterate
on steps 1-3 until the excess demands satisfy the criteria $\left\|\Delta^{g}\right\|_{\infty} \leq \epsilon$.
Once the equilibrium singles rates of men and women and of all types have been found, equations (8) and (9) along with (D1), can be used to obtain the equilibrium marriages rates, including to non-UK-born partners.

This thus provides us with the model-predicted equilibrium choice frequencies $\mu_{x^{\prime} \mid x}^{g, k}(\widehat{\boldsymbol{\theta}})$, for $k=m, f, x \in X$ and $x^{\prime} \in X \cup\{0,-1\}$ given the trial vector $\widehat{\boldsymbol{\theta}}$. Repeating for all $g \in G$ gives the full set of model-implied equilibrium choice frequencies at $\widehat{\boldsymbol{\theta}}$.

## The Likelihood Function

Let $M^{g, k}(x)$ be the observed number of individuals of gender $k=m, f$ from region $g \in G$ of type $x \in X$, and recall that $\widetilde{\mu}_{x^{\prime} \mid x}^{g, k}$ for $x^{\prime} \in X \cup\{0,-1\}$ are the empirical choice frequencies for this observed set of individuals. The likelihood contribution of this region, at the parameter vector $\widehat{\boldsymbol{\theta}}$, is hence given by

$$
\begin{equation*}
L_{x}^{g, k}(\widehat{\boldsymbol{\theta}})=\prod_{x^{\prime} \in X \cup\{0,-1\}}\left[\mu_{x^{\prime} \mid x}^{g, k}(\widehat{\boldsymbol{\theta}})\right]^{M^{g, k}(x) \widetilde{\mu}_{x^{\prime} \mid x}^{g, k}} \tag{D5}
\end{equation*}
$$

The overall $\log$-likelihood function is obtained by summing $\log L_{x}^{g, k}(\widehat{\boldsymbol{\theta}})$ over regions, gender, types, to obtain

$$
\begin{equation*}
\log L(\widehat{\boldsymbol{\theta}})=\sum_{g \in G} \sum_{k=m, f} \sum_{x \in X} \sum_{x^{\prime} \in X \cup\{0,-1\}} M^{g, k}(x) \widetilde{\mu}_{x^{\prime} \mid x}^{g, k} \log \mu_{x^{\prime} \mid x}^{g, k}(\widehat{\boldsymbol{\theta}}) . \tag{D6}
\end{equation*}
$$

The ML-estimator is hence $\widehat{\boldsymbol{\theta}}_{M L}=\arg \max _{\hat{\boldsymbol{\theta}} \in \boldsymbol{\Theta}}(\log L(\widehat{\boldsymbol{\theta}}))$. Regularity conditions and standard arguments implies that $\widehat{\boldsymbol{\theta}}_{M L} \xrightarrow{p} \boldsymbol{\theta}$ and that $\widehat{\boldsymbol{\theta}}_{M L}$ is asymptotically normal distributed. Reported standard errors are based on the estimated asymptotic variance-covariance matrix of $\widehat{\boldsymbol{\theta}}_{M L}$, approximated, in standard ways, using the Hessian matrix of the log-likelihood function.

## Appendix E: Further Estimation Results

## Aggregate Fit

Panels A and B of Figure E. 1 show the model-predicted versions of Figures 2 and 3.
Figure E.1: Predicted Distribution of Marital Status and Partner Type by Own Type and Gender
(a) Marital Status

(b) Partner Type


Notes: Predictions based on the preferred model specification (iii) in Table A.1.

## Robustness to Sample Selection

Here we report the results from re-estimating the model using two alternative samples. In case (i) we reduce the age threshold for inclusion in the sample from 25 to 20 . In case (ii) we eliminate London region from the sample.

Table E. 1 presents the estimated principal preferences in case (i). In the first, case the estimated $\Sigma$-matrix now has slightly lower values as a rule, reflecting that we include more young people who are more often single. There is also some suggestion that mixed marriages have less negative values, and thus of slightly lower ethnic complementarity, but the overall structure of the matrix is nevertheless unchanged.

Table E.1: Estimates of Principal Joint Marital Utility by Type Profile with an Age 20 Sample Threshold

|  |  | Female Type |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | White, Low | White, High | Black, Low | Black, High | Asian, Low | Asian, High |
|  | White, Low | $\begin{gathered} 0.67 * * * \\ (0.09) \end{gathered}$ | $\begin{gathered} 0.53^{* * *} \\ (0.11) \end{gathered}$ | $\begin{gathered} -4.31^{* * *} \\ (0.51) \end{gathered}$ | $\begin{gathered} -4.23^{* * *} \\ (0.50) \end{gathered}$ | $\begin{gathered} -3.78^{* * *} \\ (0.52) \end{gathered}$ | $\begin{gathered} -4.26^{* * *} \\ (0.53) \end{gathered}$ |
|  | White, High | $\begin{gathered} 0.09 \\ (0.13) \end{gathered}$ | $\begin{gathered} 1.43^{* * *} \\ (0.11) \end{gathered}$ | $\begin{gathered} -4.43^{* * *} \\ (0.53) \end{gathered}$ | $\begin{gathered} -3.22^{* * *} \\ (0.40) \end{gathered}$ | $\begin{gathered} -4.20^{* * *} \\ (0.57) \end{gathered}$ | $\begin{gathered} -2.27^{* * *} \\ (0.34) \end{gathered}$ |
| Male | Black, Low | $\begin{gathered} \hline-3.33^{* * *} \\ (0.42) \end{gathered}$ | $\begin{gathered} -3.27^{* * *} \\ (0.45) \\ \hline \end{gathered}$ | $\begin{gathered} \hline-2.35^{* * *} \\ (0.32) \end{gathered}$ | $\begin{gathered} -2.35^{* * *} \\ (0.32) \end{gathered}$ | $\begin{gathered} \hline-5.22^{* * *} \\ (0.98) \\ \hline \end{gathered}$ | $\begin{gathered} \hline-4.68^{* * *} \\ (0.70) \end{gathered}$ |
| Type | Black, High | $\begin{gathered} -3.16^{* * *} \\ (0.41) \end{gathered}$ | $\begin{gathered} \hline-2.76^{* * *} \\ (0.40) \end{gathered}$ | $\begin{gathered} \hline-2.46^{* * *} \\ (0.34) \end{gathered}$ | $\begin{gathered} -1.07^{* * *} \\ (0.20) \end{gathered}$ | $\begin{gathered} -36.2 \\ (>200) \end{gathered}$ | $\begin{gathered} -3.96^{* * *} \\ (0.59) \end{gathered}$ |
|  | Asian, Low | $\begin{gathered} -4.07^{* * *} \\ (0.50) \end{gathered}$ | $\begin{gathered} \hline-3.80^{* * *} \\ (0.51) \end{gathered}$ | $\begin{gathered} -5.44^{* * *} \\ (0.94) \end{gathered}$ | $\begin{gathered} -36.2 \\ (>200) \end{gathered}$ | $\begin{gathered} -0.20 \\ (0.19) \end{gathered}$ | $\begin{gathered} \hline-0.87^{* * *} \\ (0.22) \end{gathered}$ |
|  | Asian, High | $\begin{gathered} \hline-4.82^{* * *} \\ (0.57) \\ \hline \end{gathered}$ | $\begin{gathered} \hline-2.76^{* * *} \\ (0.39) \\ \hline \hline \end{gathered}$ | $\begin{gathered} -37.6 \\ (>200) \\ \hline \end{gathered}$ | $\begin{gathered} -38.6 \\ (>200) \\ \hline \end{gathered}$ | $\begin{gathered} -0.82^{* * *} \\ (0.23) \\ \hline \end{gathered}$ | $\begin{gathered} 0.16 \\ (0.13) \\ \hline \end{gathered}$ |

[^23]Table E. 2 presents the estimated principal preferences in case (ii). In this case many of the estimated surplus terms are numerically larger (more negative). This reflects that, in this estimation, the marriage peer effect were estimated to be lower (in the order of 0.25 rather than
0.5 , but also more imprecise). Nevertheless, the qualitative complementarity patterns remain unchanged.

Table E.2: Estimates of Principal Joint Marital Utility by Type Profile Excluding London

|  | Female Type |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | White, | White, | Black, | Black, | Asian, | Asian, |  |
|  | Low | High | Low | High | Low | High |  |
|  | White, | 0.15 | -0.11 | $-7.34^{* * *}$ | $-7.26^{* * *}$ | $-6.36^{* * *}$ | $-7.01^{* * *}$ |
| Low | $(0.12)$ | $(0.17)$ | $(0.69)$ | $(0.69)$ | $(0.68)$ | $(0.71)$ |  |
| Male | -0.77 | $1.07^{* * *}$ | $-7.32^{* * *}$ | $-5.99^{* * *}$ | $-7.70^{* * *}$ | $-4.32^{* * *}$ |  |
| White, | $(0.20)$ | $(0.09)$ | $(0.70)$ | $(0.60)$ | $(0.80)$ | $(0.51)$ |  |
| Type | Llack, | $-5.48^{* * *}$ | $-5.18^{* * *}$ | $-4.72^{* * *}$ | $-4.38^{* * *}$ | $-7.81^{* * *}$ | $-6.60^{* * *}$ |
|  | Llack, | $-5.29^{* * *}$ | $-4.94^{* * *}$ | $-5.00^{* * *}$ | $-3.27^{* * *}$ | -33.3 | $-8.15^{* * *}$ |
|  | High | $(0.54)$ | $(0.54)$ | $(0.51)$ | $(0.29)$ | $(>200)$ | $(0.97)$ |
|  | Asian, | $-6.32^{* * *}$ | $-6.08^{* * *}$ | -33.3 | -33.2 | -1.04 | $-1.66^{* *}$ |
| Low | $(0.64)$ | $(0.65)$ | $(>200)$ | $(>200)$ | $(0.28)$ | $(0.32)$ |  |
|  | Asian, | $-7.33^{* * *}$ | $-4.49^{* * *}$ | -33.6 | -34.0 | $-1.61^{* *}$ | -0.08 |
| High | $(0.72)$ | $(0.52)$ | $(>200)$ | $(>200)$ | $(0.32)$ | $(0.19)$ |  |

Notes: The sample consists of all UK-born individuals observed in the QLFS 1996-2015, born between 1965-1989, living in England, Scotland or Wales, (excluding London) aged 25 or above when observed, and with available information on gender, ethnicity, educational attainment, marital status. A person has a "low qualification" if they either no qualification or a GCSE, and they have a "high qualification" if they have an A-level qualification or higher. Standard errors in parenthesis.
${ }^{*} p<0.05,{ }^{* *} p<0.01,{ }^{* * *} p<0.001$

For completeness, Table E. 3 presents the $\Sigma$-matrix obtained from the estimation that left out the West Midlands that was used in the analysis of model fit in Section V. Comparing to Table A.2, we see that the effect of leaving out this large and ethnically diverse region had barely any effect on the estimated surplus structure.

## Appendix F: The 1990-2006 Cohorts

Here we provide details of the construction the type distribution in the "recent cohorts" by region. To this end we use the QLFS 1996-2015 and keep all individual born in the UK between 1990 and 2006. As ethnicity is directly observable, we can directly characterize the ethnic distribution by gender and region for these cohorts. However, as many of the individuals included in this sample had not completed their education by the time they were observed, we impute

Table E.3: Estimates of Principal Joint Marital Utility by Type Profile Excluding the West Midlands

|  |  | Female Type |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | White, Low | White, High | Black, Low | Black, High | Asian, Low | Asian, High |
|  | White, Low | $\begin{gathered} 0.99^{* * *} \\ (0.29) \end{gathered}$ | $\begin{gathered} 0.94^{* * *} \\ (0.26) \end{gathered}$ | $\begin{gathered} -3.65^{* * *} \\ (0.42) \end{gathered}$ | $\begin{gathered} -3.76^{* * *} \\ (0.45) \end{gathered}$ | $\begin{gathered} \hline-2.98^{* * *} \\ (0.44) \end{gathered}$ | $\begin{gathered} -3.40^{* * *} \\ (0.44) \end{gathered}$ |
|  | White, High | $\begin{gathered} 0.49 \\ (0.29) \end{gathered}$ | $\begin{gathered} 1.85^{* * *} \\ (0.39) \end{gathered}$ | $\begin{gathered} -3.62^{* * *} \\ (0.42) \end{gathered}$ | $\begin{gathered} \hline-2.53^{* * *} \\ (0.31) \end{gathered}$ | $\begin{gathered} \hline-3.37^{* * *} \\ (0.48) \end{gathered}$ | $\begin{gathered} -1.44^{* * *} \\ (0.26) \end{gathered}$ |
| Male | Black, Low | $\begin{gathered} -2.79 * * * \\ (0.39) \end{gathered}$ | $\begin{gathered} -2.61^{* * *} \\ (0.41) \end{gathered}$ | $\begin{gathered} -1.76^{* * *} \\ (0.26) \end{gathered}$ | $\begin{gathered} -1.74^{* * *} \\ (0.27) \end{gathered}$ | $\begin{gathered} -33.8 \\ (>200) \end{gathered}$ | $\begin{gathered} -4.74^{* * *} \\ (0.95) \end{gathered}$ |
| Type | Black, High | $\begin{gathered} \hline-2.58^{* * *} \\ (0.35) \end{gathered}$ | $\begin{gathered} -2.02^{* * *} \\ (0.33) \end{gathered}$ | $\begin{gathered} -1.93^{* * *} \\ (0.29) \end{gathered}$ | $\begin{gathered} \hline-0.58^{* * *} \\ (0.20) \end{gathered}$ | $\begin{gathered} -33.2 \\ (>200) \end{gathered}$ | $\begin{gathered} -3.00^{* * *} \\ (0.52) \end{gathered}$ |
|  | Asian, Low | $\begin{gathered} -3.23^{* * *} \\ (0.41) \end{gathered}$ | $\begin{gathered} -2.98^{* * *} \\ (0.43) \end{gathered}$ | $\begin{gathered} -4.34^{* * *} \\ (0.83) \end{gathered}$ | $\begin{gathered} -32.6 \\ (>200) \end{gathered}$ | $\begin{gathered} 0.35 \\ (0.29) \end{gathered}$ | $\begin{gathered} -0.30 \\ (0.27) \end{gathered}$ |
|  | Asian, High | $\begin{gathered} -3.90^{* * *} \\ (0.49) \\ \hline \end{gathered}$ | $\begin{gathered} -1.89^{* * *} \\ (0.30) \\ \hline \end{gathered}$ | $\begin{gathered} -33.5 \\ (>200) \end{gathered}$ | $\begin{gathered} -33.9 \\ (>200) \end{gathered}$ | $\begin{gathered} -0.10 \\ (0.24) \end{gathered}$ | $\begin{gathered} 0.91^{* *} \\ (0.30) \\ \hline \end{gathered}$ |

Notes: The sample consists of all UK-born individuals observed in the QLFS 1996-2015, born between 1965-1989, living in England, Scotland or Wales, (excluding the West Midlands) aged 25 or above when observed, and with available information on gender, ethnicity, educational attainment, marital status. A person has a "low qualification" if they either no qualification or a GCSE, and they have a "high qualification" if they have an A-level qualification or higher. Standard errors in parenthesis.
${ }^{*} p<0.05,{ }^{* *} p<0.01,{ }^{* * *} p<0.001$
the proportion with a high qualification by gender, region, and ethnicity in these recent cohorts.
To do so, we estimate a linear probability model for holding a high qualification (A-level+) using the estimation cohort sample observed in the QLFS 1996-2015, born in the UK and aged 25 or higher when observed. The estimated specification models the educational attainment (dummy for having an A-level+ qualification) of individual $i$, of gender $k=m, f$, living in region $g \in G$, and of ethnicity $z \in\{W, B, A\}$ and birth cohort $c_{i}$, as

$$
\begin{equation*}
q_{i k g z}=\alpha+\beta_{k}+\gamma_{g}+\delta_{z}+\kappa_{k g z} c_{i}+\varepsilon_{i k g z} \tag{F1}
\end{equation*}
$$

The model thus includes gender-, region- and ethnicity fixed-effects, and models a linear growth in rate of holding an A-level+ qualification that is also gender-, region, and ethnicityspecific. Based on the estimated model we then impute an average qualification rate by gender, region and ethnicity for the recent cohorts using the distribution of $c_{i}$ within the cell.

Table F. 1 shows aggregate type distribution for the estimating- and the recent cohorts. The table highlights an (i) increase in the rate of hold a high qualification in each ethnic group and gender, and (ii) an increase in the population share for each of the two ethnic minorities with Asian population increasing proportionately more.

Table F.1: The Distribution of Ethnicity and Qualifications in the Estimating and Recent Cohorts

|  |  | Birth Cohorts |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | 1965-1989 |  | 1990-2006 |  |
|  |  | Males | Females | Males | Females |
| White | Low | 0.458 | 0.433 | 0.173 | 0.158 |
|  | High | 0.503 | 0.524 | 0.739 | 0.754 |
|  | Total | 0.961 | 0.957 | 0.912 | 0.912 |
| Black | Low | 0.005 | 0.005 | 0.003 | 0.002 |
|  | High | 0.006 | 0.008 | 0.022 | 0.022 |
|  | Total | 0.011 | 0.013 | 0.025 | 0.024 |
| Asian | Low | 0.010 | 0.011 | 0.007 | 0.006 |
|  | High | 0.018 | 0.019 | 0.056 | 0.058 |
|  | Total | 0.028 | 0.030 | 0.063 | 0.064 |

Notes: The sample consists of all UK-born individuals observed in the QLFS 1996-2015, born in the indicated cohorts, living in England, Scotland or Wales, and with available information on gender, ethnicity, and educational attainment. A person has a "low qualification" if they either no qualification or a GCSE, and they have a "high qualification" if they have an A-level qualification or higher.

Figure F. 1 (panels a-c) shows the distribution of ethnicity across regions in the recent cohorts. Panel (d) shows the (imputed) rate of holding a high qualification by region.

Figure F.1: Demographic Composition by Region Cohorts 1990-2006


Notes: The sample consists of all UK-born individuals observed in the QLFS 1996-2015, born between 1990 and 2006 cohorts, living in England, Scotland or Wales, and with available information on ethnicity. A person has a "low qualification" if they either no qualification or a GCSE, and they have a "high qualification" if they have an A-level qualification or higher.


[^0]:    *We are very grateful to comments from Frederic Vermeulen, Monica Costa Dias and two referees, as well as from participants at the Workshop on Labour and Family Economics at York, the Society of Economics of the Household conference in Paris, the Applied Microeconomics Summer School at Lancaster, the Applied Microeconomics Summer School at Warwick, and at seminars at Royal Holloway.

[^1]:    ${ }^{1}$ There is no single agreed international definition of ethnicity and race or of the distinction between them. In general, ethnicity has many dimensions which include or combine nationality, citizenship, race, colour, language, and religion. Our broad categorization of ethnicity into White, Black and Asian people is a condensed version of the categorization used by the Office of National Statistics.

[^2]:    ${ }^{2}$ Key contributions include inter alia Galichon and Salanié (2015), Choo (2015), Dupuy and Galichon (2014), Chiappori et al. (2017), Graham (2013), and Brandt et al. (2016).

[^3]:    ${ }^{3}$ See Hansen (2000) for a comprehensive outline of post-war immigration.

[^4]:    ${ }^{4}$ We will omit Northern Ireland as the proportions of ethnic minorities there are too small for any meaningful analysis.

[^5]:    ${ }^{5}$ The GCSE is the first tier of academic qualifications in the UK, obtained at the end of the academic year in which the individual turns 16 (which also corresponds to the end of compulsory education for the cohort in question). The A-level degree is obtained at the age of 18 - after two years of post-compulsory schooling - and is the standard qualification for entry to university.
    ${ }^{6}$ Below we will also check on the sensitivity of our results to this cutoff age.

[^6]:    ${ }^{7}$ As expected, the qualification rate is also generally higher among females than among males and higher among the ethnic minorities than among Whites. These patterns are in line with findings from the literature, for instance Burgess (2014).
    ${ }^{8}$ Our classification is based on current marital status as, for married individuals, we need to measure the characteristics of their partners. This means that we classify divorced, separated and widowed individuals as single.

[^7]:    ${ }^{9}$ For this we restrict the sample to those married to UK-born partners for two reasons. First, doing so directly ties in with the modelling approach below where we model the supply of UK-born prospective partners, but not the non-UK-born supply. Second, doing so avoids having to classify in particular the qualifications of the non-UK-born partners.
    ${ }^{10}$ Note that this is conditional on the partner also being UK-born.

[^8]:    ${ }^{11}$ In a few regions we observe no intra-ethnic marriages - only a small number of inter-ethnic marriages to White partners - due to low prevalence of the ethnic minority in question. This applies to Black males in the North East and in Wales, and to Black women in Scotland and Wales.

[^9]:    ${ }^{12}$ Note that the individuals used to create the instrument can be either UK- or non-UK-born. This way we

[^10]:    capture the first generation immigrants as potential parents to the UK-born estimating sample cohorts.

[^11]:    ${ }^{13}$ Gender symmetry and the assumption that everyone marries someone of the opposite gender, means that we here assume that the male and female populations are equally large.
    ${ }^{14}$ It is otherwise common, and completely equivalent, to use a population measure notation.

[^12]:    ${ }^{15}$ We are assuming that social preferences enter utility in log form; this is contrary to Brock and Durlauf (2001). In principle, there is no obvious economic ex ante justification for either the linear or log specification. The log specification allows a semi-closed form solution.

[^13]:    ${ }^{16}$ Specifically, $\mu_{x^{\prime \prime} \mid x^{\prime}}^{m}=\mu_{x^{\prime \prime} \mid x^{\prime}}^{f}$ for any $x^{\prime}, x^{\prime \prime} \in X$.
    ${ }^{17}$ Existence and uniqueness follows trivially since the left hand side is strictly increasing in $\mu_{x_{1} \mid x_{2}}$ over [0, 1], and limits to 0 and $\infty$ at 0 and 1 respectively, and whereas the right hand side is positive constant.

[^14]:    ${ }^{18}$ Specifically, consider a sequence of economies indexed by time $t=0,1,2, \ldots, T$, where $T$ could be infinite, and let $\left\{r_{t}\right\}_{t=0}^{T}$ be the corresponding sequence of relative type supplies (while $\Delta$ and $\phi$ remain constant). Assume that for the initial cohort, $t=0$, the reference behaviour is the own cohort behaviour (i.e. the social norms are "instantaneous"), but for all subsequent cohorts $t \geq 1$, the reference behaviour is the behaviour of the previous cohort, $t-1$. It is easy to show that the sequence of equilibria in this case will be identically the same as the sequence of equilibria that would obtain if all cohorts formed instantaneous social norms: $\frac{\mu_{x^{\prime \prime} \mid x^{\prime}}^{t}}{\mu_{x^{\prime} \mid x^{\prime}}^{t}} \frac{\mu_{x^{\prime} \mid x^{\prime \prime}}^{t}}{\mu_{x^{\prime \prime} \mid x^{\prime \prime}}^{t}}=$ $\exp \left(-\frac{\Delta}{2(1+\phi)}\right)$ will emerge endogenously in the initial cohort, and will then be replicated across all subsequent cohorts.

[^15]:    ${ }^{19}$ Note that individuals in this economy do not hold preferences over specific individuals of the opposite gender, only over their types. This assumption on preferences effectively rules out sorting on any unobserved personal characteristics (Galichon and Salanié, 2015).

[^16]:    ${ }^{20}$ The online appendix also contains details of how the identification of the model extends to the case of more than two types and to more than two regions, including region fixed-effects. Note also that the additive region-fixed-effects cancel out in (10) and hence the specification tests based on this equation remains valid.

[^17]:    ${ }^{21}$ Using that, in equilibrium, $\mu_{x^{\prime \prime} \mid x^{\prime}}^{g, m}=\left(h\left(x^{\prime \prime}, f, g\right) / h\left(x^{\prime}, m, g\right)\right) \mu_{x^{\prime} \mid x^{\prime \prime}}^{g, f}$ it immediately follows that $\log \left(\zeta_{x^{\prime} x^{\prime \prime}}^{g, m}\right)=$ $\log \left(\zeta_{x^{\prime} x^{\prime \prime}}^{g, f}\right)$.

[^18]:    ${ }^{22}$ As outlined in the Appendix, we leave out one dimension of $h(\cdot, \cdot \mid g)$ to ensure that the estimated coefficients are independent. We further use bootstrapping which resamples the data to account for the sampling variation in $\log \left(\widetilde{\zeta}_{x^{\prime} x^{\prime \prime}}^{g, k}\right)$.
    ${ }^{23}$ See Chen et al. (2019) for a discussion of algorithms for ML estimation of matching equilibria under existence and uniqueness.
    ${ }^{24}$ Further details about the algorithm used to solve for the regional equilibria are presented in the online Appendix.

[^19]:    ${ }^{25}$ To be more precise, given that the scale of the preference parameters is not identified, the latter test rejects the constraint that all four social preference parameters are equal.

[^20]:    ${ }^{26}$ For ease of comparison, the figure uses the same scale in each panel. This means that a few the marital utility terms for marriages between Blacks and Asian are "below the scale" and hence not shown.

[^21]:    ${ }^{27}$ The estimated primitive preference parameters $\Sigma$ from these two cases are available in Tables E. 1 and E. 2 in the online Appendix. For completeness, we there also show the estimated $\Sigma$ from the estimation where the West Midlands was left out, used above for the model fit analysis.

[^22]:    ${ }^{28}$ Noting the some of these recent cohorts have yet to complete their education, we impute predicted distributions of completed education by gender, ethnicity and region. Details of this are provided in the online Appendix.

[^23]:    Notes: The sample consists of all UK-born individuals observed in the QLFS 1996-2015, born between 1965-1989, living in England, Scotland or Wales, aged 20 or above when observed, and with available information on gender, ethnicity, educational attainment, marital status. A person has a "low qualification" if they either no qualification or a GCSE, and they have a "high qualification" if they have an A-level qualification or higher. Standard errors in parenthesis.

    * $p<0.05,{ }^{* *} p<0.01,{ }^{* * *} p<0.001$

