IZA Institute of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17956

Sorting in the Marriage Market: A New Approach to Measuring Assortative Mating

Jeongwon Choi
Jinyoung Kim

JUNE 2025

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17956

Sorting in the Marriage Market: A New Approach to Measuring Assortative Mating

Jeongwon Choi
Korea University

Jinyoung Kim
Korea University and IZA

JUNE 2025

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Sorting in the Marriage Market: A New Approach to Measuring Assortative Mating

This paper introduces a new framework for measuring the extent of positive assortative mating (PAM) in the marriage market by relaxing the standard assumption of dichotomous sorting levels. Conventional PAM measures treat marriage sorting as a binary outcome—either perfectly matched or not—thereby failing to capture degrees of similarity between partner types. We propose a continuous measure of sorting based on trait similarity, where individuals are hypothesized to select mates according to the relative closeness of traits, which influences marital payoffs. Trait similarity is quantified using multidimensional attribute data and incorporated into a similarity-weighted matching matrix. We adapt conventional PAM indices—including the normalized trace, aggregate likelihood ratio, and perfect-random normalization—to this similarity-weighted framework. Applying our method to U.S. data on occupational, religious, and educational matching, we uncover patterns in assortative mating that are obscured under traditional approaches.

JEL Classification: J12, J24, J62

Keywords: assortative mating, partner trait, trait attributes, measure for positive assortative mating

Corresponding author:
Jinyoung Kim
Department of Economics
Korea University
145 Anam-ro Seongbuk-ku
Seoul
South Korea

E-mail: jinykim@korea.ac.kr

I Introduction

Assortative mating in the marriage market has long attracted scholarly attention due to its profound implications for family formation, income inequality, and the intergenerational transmission of human capital. When individuals select partners with similar characteristics or social status—a practice known as positive assortative mating (PAM)—they reinforce their socioeconomic positions. If the human capital of both spouses exhibits supermodularity or complementarity in raising children and developing their human capital, higher levels of PAM can exacerbate income inequality in subsequent generations (Becker, 1981).

Despite extensive research, there remains considerable debate about the evolution of assortative mating in the U.S. and other countries. Some studies have documented an increase in PAM by educational attainment or income over time in various societies (Chiappori, Salanié, and Weiss, 2017; Greenwood et al., 2014; Guell et al., 2015; Shen, 2019), while others report declining or stable trends (Eika, Mogstad, and Zafar, 2019; Gihleb and Lang, 2020).

This lack of agreement stems from the difficulty of accurately quantifying the extent of PAM. While its presence can be identified relatively easily by noting whether individuals are more likely to marry those with similar characteristics compared to what would occur under random matching, precise measurement remains challenging. Consider a scenario where each man and woman possesses a trait with one of two types—college graduates (C) or high school graduates or less (HSL). The matching between men and women, assuming equal population sizes (denoted by A), is illustrated in the following matching table.

Table 1: Matching table

	Women: C	Women: HSL
Men: C	A_{11}	A_{12}
Men: HSL	A_{21}	A_{22}

Here, A_{11}, A_{12}, A_{21}, and A_{22} represent the number of matches between individuals with the corresponding trait types, and $A_{11} + A_{12} + A_{21} + A_{22} = A$.

The table demonstrates **positive assortative mating (PAM)** if the number of couples with the same education level exceeds what would be expected under random matching. Specifically, PAM exists if and only if:

$$A_{11} > A \cdot \left[\frac{(A_{11}+A_{12})(A_{11}+A_{21})}{A^2}\right] \quad (1)$$

where the right-hand side represents the expected number of couples under random matching in which both spouses have a college education. Equivalently, the condition for PAM can also be expressed as:

$$A_{11}A_{22} > A_{12}A_{21}. \quad (2)$$

Based on this definition, most conventional measures quantify PAM by assessing how much the left-hand side exceeds the right-hand side in equation (1) or (2). For example, measures using the correlation between spouses' education levels, the χ^2 statistic, and the perfect-random normalization method evaluate the difference between $A_{11}A_{22}$ and $A_{12}A_{21}$. Alternatively, measures using the odds ratio and the like focus on the ratio of these terms. Meanwhile, those using the likelihood ratio and the normalized trace method utilize the sum of matching types $(A_{11} + A_{22})$ to measure the degree of PAM.[1] Variations in scaling factors across these measures lead to differences in observed PAM levels.

An implicit assumption underlying all these measures concerns the *degree* of positive sorting between spouses of different types. To illustrate, consider a scenario where the trait has three types: doctoral or master's degrees (DM), bachelor's degrees (BA), and high school or less

[1] The correlation between spouses' education levels or its squared value (equivalent to χ^2) has been used in studies such as Greenwood, Guner, and Knowles (2003) and Greenwood et al. (2014). The perfect-random normalization method, employed by Liu and Lu (2006) and Shen (2020), is equivalent to the minimum distance approach used by Fernández and Rogerson (2001), Abbott et al. (2019), and Wu and Zhang (2021) in the case of a two-type trait. The odds ratio is applied in Siow (2015), Chiappori, Salanié, and Weiss (2017), Chiappori et al. (2020), and Ciscato and Weber (2020), while the likelihood ratio is used by Eika, Mogstad, and Zafar (2019). The normalized trace method appears in Cheremukhin, Restrepo-Echavarria, and Tutino (2024). A comprehensive discussion of these measures is provided in Chiappori, Costa Dias, Meghir, and Zhang (2025). Some of the measures, particularly those applicable to multi-type matching, are described in Section II.

(HSL). Across all measures, it is assumed that a match between spouses with the same educational level is perfectly positive assortative (with a normalized positive sorting value of 1), while a match between spouses with different educational levels is completely non-positive assortative (with a normalized sorting value of 0). This assumption treats marriage sorting as a binary outcome, where couples are either perfectly positively sorted or completely non-positively sorted.

Under this framework, each couple with the same educational level contributes equally to the level of PAM. On the other hand, all couples with different educational levels contribute nothing, regardless of whether the match is DM-BA, DM-HSL, or BA-HSL.[2]

When a trait has only two types (as illustrated in Table 1), the assumption of binary sorting levels is harmless because we can normalize sorting levels without loss of generality. However, it can lead to inaccuracies when applied to multi-type traits. For instance, the degree of positive sorting for a match between a man with a doctoral degree and a woman with a bachelor's degree is treated the same as a match between a man with a doctoral degree and a woman with no education—both matches are assigned a sorting value of 0. Similarly, a match between two individuals with high school education is considered as sorted as a match between a high school graduate and someone with no formal education, as both matches are assigned a sorting value of 1.

This binary framework can also yield counterintuitive conclusions for other cases of spousal traits. For example, a match between a medical doctor and a trucker would be considered as sorted as a match between a medical doctor and a dentist, as long as these three are classified as distinct occupations. Likewise, a couple consisting of a Catholic and a Protestant would be treated as equally sorted as a couple consisting of a Catholic and a Buddhist.

More specifically, the assumption of dichotomous sorting levels presents the following issues:

1. **Inaccuracy in Measuring Positive Assortative Mating (PAM)**: A PhD-college couple is arguably more positively sorted than a PhD-no education couple. However, PAM measures based on dichotomous sorting levels fail to capture such distinctions, leading to inaccuracies

[2] The odds ratio differs from other measures. When applied to a two-type trait, as shown in Table 1, each couple with different trait types contributes equally and negatively to the level of PAM.

in PAM measurement. For example, if more marriages occur between doctors and dentists this year at the expense of fewer marriages between doctors and truckers, many conventional measures might indicate no change in PAM from the previous year, even though positive assortative mating has increased.

2. **Sensitivity to Trait Grouping**: Conventional PAM measures that rely on dichotomous sorting levels are highly sensitive to how individuals are grouped. Consider again the educational scenario with three categories: DM (doctoral or master's degrees), BA (bachelor's degrees), and HSL (high school graduates or less). Suppose that, due to data limitations, DM and BA are merged into a single "college graduates" category. In this case, a DM-BA match, which would have previously been treated as a non-positive sorting, is now considered perfectly positive sorting. This kind of regrouping can artificially inflate or deflate PAM values, leading to distortions in conventional measures (Gihleb and Lang, 2020).[3]

3. **Incompatibility with Multi-Type Matching**: When a trait has multiple types, dichotomous sorting levels complicate the application of conventional PAM measures. With only two sorting levels, it becomes difficult to make coherent comparisons. For instance, the difference between a PhD-PhD couple and a PhD-college couple cannot be meaningfully compared to that between a PhD-PhD couple and a PhD-high school couple. Conventional metrics, such as the odds ratio, often fail to accurately capture these nuances, making comparisons infeasible and leading to potentially misleading interpretations of PAM trends.

4. **Inconsistency with Marriage Theory**: The canonical theory of marriage (Becker, 1981) posits that marriage sorting is determined by the outputs of all possible matches, as defined by the marriage payoff matrix. Within this framework, assigning the same sorting value (0) to both a doctor-dentist match and a doctor-trucker match implies that the marriage payoffs from these two pairings are identical—an unrealistic assumption.[4] Nevertheless, conventional measures treat both matches as equally unsorted, highlighting their limitations.

---

[3] Gihleb and Lang report that PAM estimates are highly sensitive to how educational groups are defined. In the U.S., homogamy appears to have increased when all college graduates are grouped together but declined when college graduates are distinguished from those with advanced degrees.

[4] Chiappori, Costa Dias, and Meghir (2020) demonstrate that, under a model with the separable extreme value assumption for a 3-by-3 matching table, the observed matching frequency for each cell corresponds to the exponential of its marriage payoff.

5. **Dynamic Nature of Sorting**: The marriage payoffs for couples of different types—such as a college graduate and a high school graduate—can vary over time and across countries due to differences in the marriage market environment and shifts in the amount of human capital associated with different educational levels. As a result, the degree of sorting in a college-high school match may evolve over time and differ across societies. Conventional measures, which assume a fixed degree of sorting, may fail to capture these dynamics, leading to a distorted representation of the evolution of PAM within a society.

This paper introduces a novel approach to measuring the extent of positive assortative mating (PAM) while relaxing the assumption of dichotomous sorting levels. We propose that individuals select mates based on the relative closeness or similarity of different types within a given partner trait, as this similarity directly influences marriage payoffs. This influence can be either positive or negative, determining whether assortative mating is positive or negative.

Consider a scenario where assortative mating by a particular trait is positive. Two women with the same trait type may prefer partners whose trait is most similar to their own. However, due to individual resource constraints and market frictions, they may ultimately choose partners with a different trait type. Accordingly, we estimate the degree of PAM which reflects the extent of similarity between trait types, treating marriage sorting as a continuous rather than a dichotomous variable.

We derive the level of similarity between different trait types from the multidimensional attributes of the trait type. For example, in analyzing occupational sorting between partners, we utilize context-specific occupational attributes obtained from the O*NET database to estimate the similarity between two jobs (see Section III for details).

To quantify the degree of PAM, we first construct a weighted matching table, where the weight matrix consists of estimated similarity scores between trait types. We then apply conventional methods, such as the normalized trace index, the aggregate likelihood ratio, and the perfect-random normalization,[5] to this weighted matching table to calculate the degree of PAM. This paper implements this coherent framework for measuring PAM across diverse traits, including

[5] Among conventional indices, those besides these three are not readily applicable to multi-type matching (Chiappori, Costa Dias, Meghir, and Zhang, 2025).

occupation, religion, and educational attainment.

The remainder of this paper is organized as follows. Section II describes our basic framework and explains the construction procedure of our measure. Section III presents the data and methodology used to measure occupational similarity and provides empirical findings on trends in PAM in the U.S. marriage market. Sections IV and V extend our framework to sorting by religion and educational attainment, respectively. Finally, Section VI concludes with a summary of key findings and a discussion of the advantages of our approach.

II Basic Framework

Consider a society with an equal number, A, of men and women, abstracting from singles. Our analysis focuses on a single trait—partner occupation—and assumes, for simplicity, that there are three occupational types: doctor, dentist, and trucker. The observed matching pattern is represented by the following matching table, M:

Table 2: Matching table with three occupational types

M =

	Women: Doctor (j=1)	Women: Dentist (j=2)	Women: Trucker (j=3)
Men: Doctor (i=1)	a_{11}	a_{12}	a_{13}
Men: Dentist (i=2)	a_{21}	a_{22}	a_{23}
Men: Trucker (i=3)	a_{31}	a_{32}	a_{33}

where a_{ij} represents the proportion of matches between men in occupation i and women in occupation j, calculated as the number of matches divided by the total population size A. We impose the constraints:

$$a_{ij} \geq 0, \ \sum_{i=1}^{3} \sum_{j=1}^{3} a_{ij} = 1 \tag{3}$$

To quantify occupational similarity, we define the degree of similarity between occupations i and j as s_{ij}, derived from comparing their attributes (see Sections III-V for details on the derivation procedure). We normalize s_{ij} within the range [0, 1], where:

s_{ij} = 1 indicates that occupations i and j share identical attributes, and

s_{ij} = 0 represents occupations that are entirely dissimilar.

The similarity matrix *S*, which represents the degree of assortative matching in our framework, is defined as follows:

Table 3: Similarity matrix with three occupational types

S =

	Women: Doctor (*j*=1)	Women: Dentist (*j*=2)	Women: Trucker (*j*=3)
Men: Doctor (*i*=1)	s_{11}	s_{12}	s_{13}
Men: Dentist (*i*=2)	s_{21}	s_{22}	s_{23}
Men: Trucker (*i*=3)	s_{31}	s_{32}	s_{33}

Several important considerations arise regarding the similarity matrix in general. First, the similarity score is not necessarily symmetric ($s_{ij} \neq s_{ji}$), as occupational attributes can differ by gender. For instance, the tasks and required skills of a male nurse may not be identical to those of a female nurse, leading to variations in occupational similarity scores depending on gender.[6]

Second, conventional PAM measures impose restrictive assumptions by treating the similarity matrix as an identity matrix, where s_{ii} = 1 for all i and s_{ij} = 0 for all $i \neq j$. This approach

[6] The occupational similarity scores may also vary by gender because, for example, the tasks and required skills of a male nurse may not be identical to those of a female nurse. However, the gender difference is not incorporated in our occupational similarity analysis due to lack of data (see Section III), but in the educational similarity analysis (see Section V).

assumes that partners in the same occupation always have the highest similarity, while those in different occupations have none. However, this dichotomous framework overlooks the fact that similarity exists on a continuum, with some occupations being more similar than others.

Third, occupational similarity is often less than perfect, even within the same occupational category, meaning that s_{ii} may not always equal 1. For example, if medical specialties such as cardiology and neurology are merged into a single "doctor" category due to data limitations, the similarity score for doctors would reflect a weighted average of the similarity scores among these specialties. As a result, s_{ii} may be less than 1. However, when detailed occupational data are unavailable, it is often necessary to set s_{ii} = 1 for practical reasons.

Finally, similarity scores can evolve over time as occupational attributes shift, altering the relationships between different occupations. This means that changes in the degree of PAM may result not only from shifts in the matching matrix *M* but also from changes in the similarity matrix *S*, reflecting broader labor market dynamics.

Using the matching table *M* and the similarity matrix *S* as key building blocks, we apply conventional methods—normalized trace, aggregate likelihood ratio, and perfect-random normalization—to derive various versions of our PAM measure. Chiappori, Costa Dias, Meghir, and Zhang (2025) discuss these methods in the context of a similarity matrix defined as an identity matrix.

Normalized trace and Weighted similarity: The normalized trace is based solely on the diagonal elements of the matching table *M*, disregarding off-diagonal elements (Cheremukhin, Restrepo-Echavarria, and Tutino, 2024). In our framework, it is defined as the sum of the products of the diagonal elements from *M* and *S*:

$$NT = a_{11}s_{11} + a_{22}s_{22} + a_{33}s_{33}. \tag{4}$$

This measure captures the average degree of assortative mating, weighted by the frequencies of actual pairings represented in the matching table *M*, but considers only the diagonal elements. Because it ignores off-diagonal elements of the similarity matrix, it overlooks the correlation between trait types and fails to incorporate valuable information on cross-type assortativeness.

To address this limitation, we can extend the measure to incorporate all elements by applying the Frobenius inner product to M and S:

$$WS = \sum_{i,j} a_{ij} \cdot s_{ij}. \tag{5}$$

We refer to this measure as **weighted similarity** (***WS***), which presents the degree of assortative matching across the entire sample, with the weights corresponding to the frequency of pairings in the matching table M.

A key limitation of both measures is their inability to isolate the effect of changes in marginal distributions of a partner trait for men and women. For instance, an increase in the number of female college graduates would mechanically lead to more matches between male and female college graduates, thereby inflating the measured level of assortativeness in these indices, even if underlying sorting patterns remain unchanged.

In conventional PAM measurement, where s_{ii} = 1 for all i and s_{ij} = 0 for all $i \neq j$, both the normalized trace and weighted similarity reduce to:

$$NT = WS = a_{11} + a_{22} + a_{33}. \tag{6}$$

Aggregate likelihood ratio: The type-specific likelihood ratio LR_{ij} measures marital sorting between men of type i and women of type j (Eika, Mogstad, and Zafar, 2019):

$$LR_{ij} = \frac{a_{ij}}{r_{ij}}, \tag{7}$$

where r_{ij} is the probability of random matching between men of type i and women of type j, calculated as:

$$r_{ij} = \sum_{k=1}^{3} a_{ik} \sum_{k=1}^{3} a_{kj}. \tag{8}$$

Since the similarity score cancels out in both the numerator and denominator, it does not appear in LR_{ij}.

The aggregate likelihood ratio is a weighted average of the type-specific likelihood ratios, where the weight on each type-specific likelihood ratio in our framework corresponds to the similarity intensity share of matches under random matching:

$$LR = \sum_{i,j} \frac{r_{ij} \cdot s_{ij}}{\sum_{i,j} r_{ij} \cdot s_{ij}} \cdot \frac{a_{ij}}{r_{ij}} = \frac{\sum_{i,j} a_{ij} \cdot s_{ij}}{\sum_{i,j} r_{ij} \cdot s_{ij}} \tag{9}$$

On the right-hand side of the last equality, the numerator equals the weighted similarity (*WS*), which represents the average degree of assortative mating, weighted by the frequencies of actual pairings represented in the matching table *M*. The denominator is the weighted similarity based on counterfactual match proportions under random matching. Thus, this ratio quantifies the magnitude of similarity in observed matches relative to that in random matches.

In conventional PAM measurement, where $s_{ii} = 1$ for all i and $s_{ij} = 0$ for all $i \neq j$, the aggregate likelihood ratio simplifies to:

$$LR = \frac{a_{11} + a_{22} + a_{33}}{r_{11} + r_{22} + r_{33}}. \tag{10}$$

This measure is utilized in Greenwood et al. (2014).

Perfect-random normalization: Perfect-random normalization is designed to determine where the observed matching distribution falls between two extremes: random matching and perfectly assortative matching. This method allows us to measure the extent of assortative matching relative to these two benchmarks (Liu and Lu, 2006; Shen, 2020).

We define the perfect-random normalization measure as follows:

$$PR = \frac{\sum_{i,j} a_{ij} \cdot s_{ij} - \sum_{i,j} r_{ij} \cdot s_{ij}}{\sum_{i,j} p_{ij} \cdot s_{ij} - \sum_{i,j} r_{ij} \cdot s_{ij}} \tag{11}$$

where p_{ij} represents the proportion of matches between men of type i and women of type j under perfectly positive assortative matching, which we define below.

The numerator of this equation captures the difference between the observed matching pattern and the random matching benchmark, while the denominator represents the total possible variation between the random and perfectly assortative matching scenarios. This ratio therefore indicates how much of the maximal possible variation in PAM is reflected in the observed data.

Perfectly assortative matching is defined as the matrix of counterfactual matching proportions $[p_{ij}]$ that maximizes

$$\sum_{i,j} p_{ij} \cdot s_{ij} \tag{12}$$

subject to the constraints:

$$\sum_k p_{ik} = \sum_k a_{ik} \ \forall i, \quad \sum_k p_{kj} = \sum_k a_{kj} \ \forall j, \quad \sum_{i,j} p_{ij} = 1, \quad p_{ij} \geq 0.$$

This maximization problem follows a standard linear programming formulation and can be efficiently solved using numerical computation packages such as Matlab. The resulting matrix $[p_{ij}]$ provides an upper bound on occupational similarity in assortative mating while preserving the observed marginal distributions.

For a two-type trait, as shown in Table 1, the solution is straightforward, provided that the similarity scores satisfy the condition $s_{ii} > s_{ij}$ for all $i \neq j$. This condition is naturally met in conventional measures, where perfectly assortative matching is assumed to occur only when individuals pair with a partner of the same type (s_{ii}= 1, s_{ij}= 0 for $i \neq j$). However, even under a more flexible assumption regarding s_{ij}, the solution remains unchanged.

In a perfectly assortative matching scenario for Table 1, if the number of male college graduates (C) exceeds the number of female college graduates, then p_{11} (the probability of a male C matching with a female C) equals the proportion of female college graduates. Similarly, p_{22} (the probability of a male HSL matching with a female HSL) equals the proportion of male HSLs. Any excess male college graduates who cannot match with a female college graduate are then assigned to female HSLs.

The derivation of p_{ij} becomes more complex when a trait includes multiple types, as shown in Table 2. Under the conventional assumption that s_{ii}= 1 and s_{ij}= 0 for all $i \neq j$, the solution for perfectly assortative matching can produce counterintuitive results. For example, if the number of male doctors exceeds the number of female doctors, the assignment of unmatched male doctors to female dentists or female truckers has no effect on the degree of perfection in PAM, despite the fact that these occupations may differ significantly in similarity.

Our method for deriving p_{ij} within the similarity matrix framework addresses this issue by incorporating gradual similarity, allowing for a more intuitive definition of perfectly assortative matching. If doctors are more similar to dentists than to truckers, then in a perfectly sorted matching, we should observe more doctor-dentist pairs than doctor-trucker pairs.[7]

By employing the four methods discussed earlier, we derive four versions of our PAM measure, each of which addresses the limitations of conventional measures discussed in Section I—except for one: sensitivity to trait grouping. To illustrate this issue, consider a three-type trait, as shown in Table 2. Suppose types 1 and 2 are merged into a single type due to data limitations, and only the resulting 2-by-2 matching table and similarity matrix are available. The transformed matching table M' and similarity matrix S' are then given by:

$$M' = \begin{bmatrix} a_{11} + a_{12} + a_{21} + a_{22} & a_{13} + a_{23} \\ a_{31} + a_{32} & a_{33} \end{bmatrix}. \qquad (13)$$

$$S' = \begin{bmatrix} s'_{11} & s'_{21} \\ s'_{12} & s'_{22} \end{bmatrix},$$

When we apply the normalized trace method, our PAM measure remains invariant under this merging process if the following conditions hold:

$$s'_{22} = s_{33}, \qquad (14)$$

$$s'_{11} = \frac{a_{11}}{a_{11}+a_{12}+a_{21}+a_{22}} \cdot s_{11} + \frac{a_{22}}{a_{11}+a_{12}+a_{21}+a_{22}} \cdot s_{22}.$$

The values of s'_{11} and s'_{22} that ensure PAM invariance can be obtained if we have prior knowledge of a_{11}, a_{22}, s_{11}, s_{22}, and s_{33}. However, if any of these values is unknown, achieving an identical PAM value is not guaranteed.

When alternative methods are used instead of the normalized trace approach, the necessary conditions for PAM invariance become even more restrictive, making these methods less robust against trait grouping sensitivity.

[7] Instead of explicitly using the similarity matrix, Shen (2019) assumes that in perfectly assortative matching, individuals who cannot be paired with partners of the same education level due to group size imbalances are matched with those at the next highest level. For education, this monotonicity assumption may not be overly restrictive. However, for traits where types cannot be ranked in a strict order, this assumption becomes impractical.

III Occupational Assortative Mating: First Application

In this section, we apply our methodology to quantify positive assortative mating (PAM) within the occupational context. While most research on assortative mating has focused on educational attainment and income levels, occupational sorting remains a critical but often overlooked dimension. Since labor income is directly tied to occupation, understanding occupational assortative mating is essential for fully assessing its implications for income inequality.

Our investigation is motivated by compelling evidence of occupational assortative mating in the dataset we analyze. For instance, in 2024, the probability of a married couple sharing the same occupation was 14 times higher than that of a randomly matched pair (see Figure 1 and its accompanying note for details).

Despite the significance of occupational assortative mating, relatively few recent studies have examined this topic. Most existing research focuses on patterns observed up to the 1980s, while more recent studies often rely on highly simplified occupational classifications (Han and Qian 2021; Schwartz 2013; Schwartz et al. 2021). Additionally, previous studies have typically analyzed occupational sorting from narrow perspectives, such as prestige scores or the likelihood of spouses having the same occupation (Hayatt, 2015; Mansour and McKinnish, 2018). In contrast, our study employs a systematic framework to measure the extent of occupational assortative mating more comprehensively.

Specifically, we apply our methodology described in Section II to quantify the similarity between occupations, which we use to analyze occupational assortative mating in the United States. Our approach decomposes job characteristics into a finite number of components that capture various occupational attributes. These components reflect multiple dimensions, including cognitive, physical, and sensory abilities, as well as social, technical, problem-solving, and resource management skills. To measure the similarity between the occupational characteristics of husbands and wives, we employ conventional similarity metrics, using vectors of job trait components.

This section is structured as follows. First, we provide a detailed description of the data. Next, we outline the construction of the similarity matrix for occupations. Finally, we examine the extent and trends of occupational assortative mating in the United States over the past several decades.

3.1. Data Description

Our analysis of occupational assortative mating in the United States relies on two primary datasets. To capture the attributes associated with different occupations, we use the Occupational Information Network (O*NET) database. Additionally, we obtain individual marital and socioeconomic data from the Current Population Survey (CPS).

The O*NET database is a comprehensive source of occupational characteristics, developed in 1998 under the sponsorship of the U.S. Department of Labor. It replaced the Dictionary of Occupational Titles (DOT) by providing more accessible, up-to-date, and detailed information. The database is compiled from surveys of job incumbents and occupational analysts, covering 974 occupations classified under the Standard Occupational Classification (SOC) system.

The O*NET program recommends using data released after 2003 for longitudinal studies, as regular sampling and standardized occupational attribute assessments were introduced that year, enhancing data comparability, consistency, and reliability in trend analysis. Following this guidance, we use O*NET data from 2003 to 2024. The database has been updated quarterly since 2003, and for our analysis, we use the most recent updates available each year.

Our study focuses on six worker-oriented domains of job characteristics, comprising 148 job attributes: Abilities (AB), Occupational Interests (OI), Knowledge (KN), Skills (SK), Work Styles (WS), and Work Values (WV). These domains were selected for their relevance in capturing fundamental aspects of job performance and worker compatibility. In contrast, we exclude other O*NET domains—such as tasks, work context, or workforce characteristics—as these primarily describe environmental or structural aspects of occupations rather than intrinsic, worker-centered attributes. For example, the Work Context domain includes items such as frequency of telephone use, exposure to contaminants, or the degree of physical proximity to others—features that describe job settings rather than the traits of the individuals performing them. Similarly, Tasks often capture occupation-specific duties that are not comparable across occupational groups. Given our aim to capture latent characteristics that influence sorting in the marriage market, we restrict our analysis to domains that reflect individual capabilities, preferences, and values.

Attributes within the Abilities (AB) domain measure the extent to which an occupation requires mental, physical, eye-hand coordination, and sensory abilities for job performance. The Occupational Interests (OI) domain captures individuals' preferences for specific work activities, shaped by underlying motivations and inclinations. This domain classifies occupations based on personality types such as Realistic (hands-on, practical work), Artistic (creative, design-oriented tasks), and Enterprising (persuasive, leadership-oriented roles). The Knowledge (KN) domain reflects the breadth and depth of subject-matter expertise necessary for various occupations, covering disciplines such as business, mathematics, and communications, which define the foundational knowledge required for job performance and career specialization. The Skills (SK) domain includes both basic and cross-functional skills essential for occupational performances. Basic skills (e.g., reading comprehension, critical thinking) are fundamental across occupations, while cross-functional skills (e.g., problem-solving, coordination) enhance adaptability and interpersonal interactions. Domain Work Styles (WS) represents personality-related attributes that influence workplace behavior and job performance, such as dependability, initiative, and stress tolerance, which shape how individuals engage with tasks, colleagues, and organizational structures. Finally, the Work Values (WV) domain reflects core beliefs and preferences that drive job satisfaction and career decisions, including achievement, autonomy, recognition, and relationships.[8]

For each attribute, O*NET provides two numerical measures: importance and required level for job performance. The importance measure reflects how relevant an attribute is to a particular occupation (rated on a 1-to-5 scale). The required level measure represents the degree of proficiency or intensity needed to perform the job (rated on a 0-to-7 scale). We use only the importance measure in our analysis, as it is strongly correlated with the level measure and is available across all domains. Based on this occupational attribute data, we construct the

---

[8] Ability area comprises 52 elements and is grouped into four categories: cognitive abilities (21 elements), physical abilities (9 elements), psychomotor abilities (10 elements), and sensory abilities (12 elements). There are 6 elements in occupational interests: realistic, investigative, artistic, social, enterprising, and conventional occupations. Knowledge has 33 elements with ten subfields: arts and humanities (5 elements), business and management (6 elements), communications (2 elements), education and training (1 element), engineering and technology (5 elements), health services (2 elements), law and public safety (2 elements), manufacturing and production (2 elements), mathematics and science (7 elements), and transportation (1 element). Skills consist of 35 elements and are classified into six subgroups: basic skills (10 elements), social skills (6 elements), technical skills (11 elements), complex problem solving skills (1 element), system skills (3 elements), and resource management skills (4 elements). Work styles are measured by 16 elements, which are achievement orientation (3 elements), adjustment (3 elements), conscientiousness (3 elements), independence (1 element), interpersonal orientation (3 elements), practical intelligence (2 elements), and social influence (1 element). Finally, Work values have 6 elements: achievement, working conditions, recognition, relationships, support, and independence.

similarity matrix S (see section 3.2 for details on its construction). Table A.2.1 provides a detailed description of the attributes.

To capture individual marital and socioeconomic information, we use data from the CPS, a monthly household survey conducted jointly by the U.S. Census Bureau and the U.S. Bureau of Labor Statistics. Specifically, we utilize the March CPS dataset, supplemented by the Annual Social and Economic Supplement (ASEC). This dataset provides detailed occupation information and includes unique family identifiers, allowing us to accurately match spouses.

Our analysis uses CPS data from 2003 to 2024, focusing on married couples where at least one spouse is aged 26–60. We select 2003 as the starting point to align the CPS sample period with O*NET data recommendations, ensuring consistency in occupational measures. The age range of 26 to 60 is chosen to capture individuals who have largely completed their formal education and established their careers while excluding older individuals more likely to have exited the labor force. This approach ensures that our analysis reflects occupational sorting patterns among actively working married couples.

Based on the occupational information of spouses in CPS, we construct the matching table M. However, since the CPS occupation coding system follows the Census Occupation Classification, while the similarity matrix S constructed from O*NET is based on SOC codes, we must map occupation codes between the two systems. To accomplish this, we use the crosswalk published by the U.S. Bureau of Labor Statistics Employment Projections program.

Due to challenges in mapping, several modifications are made to the matching table. Some occupations appear in CPS but not in O*NET. Our CPS sample includes 372,348 dual-earner couples, meaning both partners report an occupation. Among these, 24,926 cases (6.7% of the sample) correspond to occupations not reported in O*NET and are thus excluded from the matching table. Conversely, some O*NET occupations have no corresponding occupations in CPS. In such cases, these occupations remain in the matching table, but their proportion is set to zero. Additionally, some CPS occupations correspond to multiple O*NET occupations. In these cases, the similarity matrix must be adjusted accordingly. The following section provides further details on these modifications.

The dimension of the modified matching table M (total number of occupations) varies by year, ranging from 453 to 530 occupations.

3.2. Estimation of Occupational Similarity

The six domains of job characteristics in O*NET encompass a total of 148 attributes. To reduce the high dimensionality of these attributes and address potential redundancies, we apply exploratory factor analysis (EFA)—a technique widely used in the literature on occupational similarity (Heckman et al., 2013; Makridis et al., 2023). Among the various dimensionality reduction methods, EFA is particularly suitable for identifying latent factors that capture variations in occupational attributes. Following standard practice, we retain only those factors with eigenvalues greater than one. This procedure yields 18 factors that collectively explain 94.66% of the total variance in the original attribute set, ensuring that the extracted factors offer a comprehensive and parsimonious representation of occupational characteristics.

To estimate pairwise similarity between occupations, we use cosine similarity as our primary metric (Okumura and Usui, 2016), applying it to vectors composed of the 18 attribute factors. This choice offers several methodological advantages. First, cosine similarity captures the direction of attribute vectors—emphasizing the relative composition of traits—rather than their magnitude. In occupational contexts, this is particularly appropriate, as the structure or combination of skills, abilities, and knowledge tends to matter more for job performance than absolute intensity levels. This rationale underlies the widespread use of cosine similarity in O*NET documentation (Dahlke et al., 2022) and in related research on occupational structure and task-based analysis (e.g., Acemoglu et al., 2022; Henning et al., 2025).

Second, in the context of spousal matching, we hypothesize that similarity in the type or structure of occupational traits—rather than the depth or intensity—may more strongly influence assortative mating. Sharing similar kinds of occupational characteristics may foster better communication, mutual understanding, or lifestyle compatibility, thereby enhancing the likelihood of forming a partnership.

Third, and importantly, cosine similarity avoids some of the pitfalls associated with magnitude-based measures such as Euclidean distance. Because absolute trait levels may rise with age or work experience, Euclidean distance may exaggerate differences between otherwise similar occupational profiles, particularly for older survey respondents. Cosine similarity, being scale-invariant, mitigates these distortions and more faithfully captures underlying similarity in occupational structure.[9]

Using cosine similarity, we construct a similarity matrix S with the same dimensionality as the occupation matching matrix M, where each element s_{ij} denotes the estimated similarity between CPS occupations i and j, as defined in Section II.

A key challenge in constructing a compatible similarity matrix S is that O*NET defines occupations at a much more granular level (approximately 1,000 occupations) than CPS, which categorizes occupations into broader groups (about 500 occupations). To address this discrepancy, we aggregate O*NET occupations into CPS occupational categories before computing similarity measures.

In cases where a single CPS occupation corresponds to multiple O*NET occupations, we compute s_{ij} by considering all possible pairwise comparisons between O*NET occupations. Specifically, for similarity between CPS occupations i and j, which map to multiple O*NET occupations, we define:

$$s_{ij} = \frac{1}{|i|\cdot|j|} \sum_{p,q} \hat{s}_{pq} \tag{15}$$

where $|i|$ and $|j|$ denote the number of O*NET occupations mapped to CPS occupations i and j, respectively, and $\hat{s}_{pq}$ represents the similarity between O*NET occupations p and q. Since employment shares at the O*NET occupational level are not available, we compute s_{ij} as the arithmetic average of all O*NET-to-O*NET similarity measures within each CPS occupation classification.

A direct consequence of this approach is that self-similarity values (s_{ii}) for some CPS occupations may not be exactly 1. When multiple O*NET occupations map to a single CPS occupation, self-similarity is calculated based on all pairwise comparisons among the corresponding O*NET occupations. Since some O*NET occupations within the same CPS

---

[9] As an alternative metric, we also employ Euclidean similarity (Robinson, 2018), which accounts for both the magnitude and composition of occupational attributes—capturing differences in both intensity and variety. In contrast, cosine similarity focuses solely on the relative composition of attributes, disregarding intensity and emphasizing proportional similarity in occupational skill profiles. For example, a surgeon and a dentist may exhibit high cosine similarity due to similarly structured skill sets, even if the intensity levels differ. By contrast, a surgeon and a truck driver would display low cosine similarity (due to divergent skill types) and low Euclidean similarity (due to differences in both type and intensity).

category have slightly different attribute compositions, the average self-similarity score may fall below 1. On average, 200 CPS occupations (ranging from a minimum of 157 to a maximum of 223) are mapped to two or more O*NET occupations throughout the analysis period. In these cases, s_{ii} is less than 1, accounting for approximately 36.3% of all occupations.

This approach may underestimate s_{ij} because we take the arithmetic mean while the share of $\hat{s}_{pp}$ is likely to be larger than that of $\hat{s}_{pq}$ $(p \neq q)$. However, since O*NET occupations mapped to the same CPS occupation are expected to have similar attributes, the distribution of marriages across these detailed occupations is likely to be relatively uniform. As an alternative, we tested a scenario in which s_{ij} is computed using the two average attribute vectors for CPS occupations i and j, with each average derived from the attribute vectors of all O*NET occupations mapped to the respective CPS category. In this case, the self-similarity values (s_{ii}) are equal to 1 by construction. This alternative specification yields qualitatively similar results to our baseline.

Since similarity values play a crucial role in measuring occupational assortative mating, it is essential to examine their distribution and variation across occupations. The annual average self-similarity measure (s_{ii}) ranges from 0.945 to 0.965 for the cosine measure. Over the entire period, the mean self-similarity values are 0.956. In terms of overall occupational similarity, the cosine similarity measure has an average of 0.502 with a standard deviation of 0.102. The minimum similarity score is 0.118 and the maximum is 1. Over the sample period, the annual average similarity values remain relatively stable, with cosine similarity fluctuating between 0.500 and 0.504. The standard deviation of similarity values also shows little variation over time, suggesting that occupational similarity patterns remain largely consistent across survey years. The distribution of occupational similarity scores is presented in Table 4.

To assess the effectiveness of our similarity measures, we identify the five occupations with the highest and lowest similarity to physicians & surgeons. Occupations with the highest similarity are: podiatrists, optometrists, veterinarians, healthcare diagnosing or treating practitioners (all others), and dentists. Occupations with the lowest similarity are: dancers and choreographers, telemarketers, actors, exercise trainers and group fitness instructors, and graders and sorters of agricultural products. As expected, the lowest cosine similarity list comprises occupations with distinct types of skills. The similarity matrix of a selected set of occupations is presented in Table 5.

3.3. Measurement of Occupational Assortative Mating

Using the matching table constructed in Section 3.1 and the similarity matrix developed in Section 3.2, we apply our PAM measurement framework to analyze trends in occupational assortative mating in the United States.

Our analysis focuses on birth cohort patterns to examine how occupational PAM has evolved across generations. Among the four measurement approaches discussed in Section II, we begin with the perfect-random normalization measure, which serves as a benchmark. This method is particularly informative because it adjusts for changes in both perfectly assortative and random matching scenarios over time, providing a properly scaled measure of PAM.

The cohort analysis is structured by the husband's year of birth, grouped into five-year intervals. The sample includes men born between 1950 and 1989, with the earliest cohort covering 1950-1954 and the most recent cohort covering 1985-1989. This range encompasses the majority of our full sample, while excluding marginal cohorts with limited representation or insufficient exposure to the labor market. Specifically, individuals born before 1950 account for only 3.0% of the sample, and those born after 1989 represent just 1.7%.

Since cohort classification is based on the husband's birth year—and the sample is restricted to couples in which at least one spouse is aged 26 to 60—these marginal cohorts are only partially represented within the 2003–2024 survey window. For example, individuals born between 1940 and 1944 would have exceeded the upper age limit by 2003, while those born between 1995 and 1999 would not reach the minimum threshold until the very end of the sample period. Although such individuals may still appear in the sample if their spouse meets the age criteria, the likelihood of inclusion is considerably lower. To enhance precision and comparability, we exclude these cohorts from the cohort-based analysis.

A key methodological challenge in the cohort-based analysis is that the occupational similarity matrix is constructed by survey year rather than by birth cohort. To address this, we generate a cohort-specific matching matrix for each year in which members of a given cohort appear in the sample, applying the corresponding year-specific similarity matrix. As a result, the cohort-level analysis is based on a set of matching matrices, with the total number equal to the number of birth cohorts multiplied by the number of survey years. This approach ensures that our estimates reflect evolving occupational structures while maintaining internal consistency within the similarity-based PAM framework.

To compute the cohort-level PAM measure, we proceed in two steps. First, we calculate year-specific weighted similarity scores for each cohort using the cohort's matching matrix and the corresponding year-specific similarity matrix. Second, we aggregate these scores across survey years, weighting each year by the proportion of the cohort observed in that year's sample. The same procedure is used to calculate the weighted similarity based on counterfactual match proportions under random matching and perfect matching.

Results by Birth Cohort

Figure 2 presents the perfect-random normalization measure of occupational PAM across birth cohorts, based on cosine similarity. Each cohort-specific PAM estimate is accompanied by a 95% confidence interval in this figure and all subsequent ones, calculated using standard errors derived from the delta method.[10]

The figure exhibits a U-shaped pattern, with lower levels of assortative mating among individuals born between 1956 and 1964, compared to higher levels among those born in the early 1950s, 1970s, and 1980s. The decline in PAM for the mid-century cohorts is notable: the measure for the 1955–1959 cohort is 5% lower than that for the 1950–1954 cohort and 12% lower than that for the 1985–1989 cohort.

Figure 3 reports results using the aggregate likelihood ratio measure. The overall pattern is qualitatively consistent with Figure 2, showing a dip for cohorts born between 1956 and 1964. However, the magnitude of the decline is more modest: the measure for the 1955–1959 cohort is only 2% lower than that for the 1985–1989 cohort.

Figure 4 displays results based on the weighted similarity measure, which follows the same

[10] In the occupational application, only the match proportions (a_{ij}) are treated as random variables in the delta method, since the similarity scores (s_{ij}) are fixed and not sample-derived. In contrast, for the religion and education applications, both a_{ij} and s_{ij} are treated as random variables.

general pattern but exhibits minimal variation across cohorts as in Figure 3. The weighted similarity for the 1955–1959 cohort is just 1% lower than that for the 1985–1989 cohort.

Figure 5 presents results using the normalized trace measure. Like the previous figures, it indicates lower PAM among mid-century cohorts and higher levels among earlier and later cohorts. The variation in this measure is more pronounced: the normalized trace for the 1955–1959 cohort is 11% lower than that for the 1985–1989 cohort.[11]

Comparison to the Conventional Approach

Appendix Figure A.1.1 shows the perfect-random normalization measure under the conventional approach, which assumes a fixed identity similarity matrix. In contrast to Figure 2, this figure reveals no meaningful trend across cohorts prior to the most recent one, which exhibits an unusually high PAM value. This comparison highlights how disregarding relative similarities between occupations can obscure important generational differences in assortative mating.

Figure A.1.2 presents the cohort trend for the aggregate likelihood ratio under the conventional approach. Its pattern closely resembles that in Figure A.1.1, with no consistent rise in PAM until the 1985–1989 cohort.

Figure A.1.3 reports results for the weighted similarity and normalized trace measures under the conventional approach. As discussed in Section II, these two measures are equivalent when the identity similarity matrix is used and are also functionally identical to the probability that spouses share the same occupation. The pattern in Figure A.1.3 closely matches that of the normalized trace under the similarity-based approach, with a 13% increase in PAM from the 1955–1959 cohort to the 1985–1989 cohort.

3.4. Discussion on Rising Occupational PAM in Recent Cohorts

What accounts for the rise in occupational positive assortative mating (PAM) among more recent cohorts in the United States? While a comprehensive causal analysis is beyond the scope

[11] Appendix A.2 reports PAM estimates based on Euclidean similarity and discusses several reasons why the resulting cohort patterns differ from those derived using cosine similarity.

of this paper, several plausible contributing factors warrant discussion.

Rising Educational Attainment among Women

Over the past several decades, women's educational attainment has increased markedly, narrowing the gender gap in education (Bureau of Labor Statistics, 2017). As a result, women and men have become more likely to pursue similar types of careers, potentially contributing to greater occupational assortative mating.

This link between educational and occupational sorting is evident in our empirical findings. The U-shaped pattern of occupational PAM observed in Figure 2 closely mirrors the cohort trend in educational PAM shown in Figure 12, both estimated using the perfect-random normalization measure.

To further assess the role of educational attainment in driving occupational assortative mating, we construct a counterfactual scenario in which the joint distribution of spousal educational attainment is held fixed at the level observed in the 1955–1959 cohort. This approach allows us to isolate the effect of shifting educational distributions on the observed increase in occupational PAM across cohorts. Specifically, we impose the 1955–1959 joint distribution on the total number of observations in the 1985–1989 cohort to derive hypothetical cell counts by spousal educational pairing.[12] If the hypothetical count is lower (higher) than the observed count in a given cell, we randomly exclude (duplicate) couples to adjust the joint educational distribution of the 1985–1989 cohort to match that of the 1955–1959 benchmark cohort.

The results show that the perfect-random normalization measure of occupational PAM for the 1985–1989 cohort falls to 0.1584 under the counterfactual scenario, compared to an observed value of 0.1610. This suggests that approximately 14% of the observed increase in occupational PAM from the 1955–1959 cohort (PAM = 0.1421) to the 1985–1989 cohort (PAM = 0.1610) can be attributed to rising educational attainment among women.[13]

[12] As in Section V later, educational attainment is categorized into four groups: less than high school, high school graduate, some college, and college degree or higher.

[13] Rising educational attainment among women accounts for 26%, 47%, and 68% of the observed increase in occupational PAM between the 1955–1959 and the 1985–1989 cohorts when measured using the aggregate likelihood ratio, the weighted similarity, and the normalized trace, respectively.

Increasing Female Labor Force Participation

Rising female labor force participation may partially account for the observed increase in occupational assortative mating, especially given certain data limitations. Because occupational attribute data from O*NET are unavailable for homemakers, we cannot include their occupations in our analysis of occupational similarity. As more women enter the labor market, a larger share of female spouses becomes represented in the occupational matching data. Depending on whether the occupations of newly participating women are more or less similar to those of their male spouses than those of previously participating women, the observed trend in occupational PAM may overstate or understate the true underlying change.

To assess the impact of increased female labor force participation, we adopt a strategy analogous to the previous subsection. Specifically, we construct a counterfactual scenario in which the age profile of women's labor force participation is held fixed at the level observed for the 1955–1959 birth cohort. Since women from this cohort are not observed at younger ages in our sample, we estimate the full age profile using the entire sample, assuming a common profile shape across cohorts but allowing for cohort-specific intercepts.

Under this counterfactual, the perfect-random normalization measure of occupational PAM for the 1985–1989 cohort increases from 0.1610 to 0.1685—an increase of 0.0075, or roughly 40 percent of the total cohort difference.[14] This result suggests that rising female labor force participation has modestly dampened the observed growth in occupational PAM: had participation remained at mid-century levels, measured PAM in recent cohorts would be even higher.

Structural Shifts in the Labor Market

Over the past several decades, the U.S. labor market has undergone substantial structural transformation, marked by the disappearance of certain occupations and the emergence of new ones. These compositional shifts can influence observed trends in occupational assortative mating. In particular, if the occupations that disappear (or emerge) are, on average, more or

[14] Similarly, the aggregate likelihood ratio increases by approximately 0.4 percent (from 1.0852 to 1.0894); the weighted similarity measure rises by about 0.6% (from 0.5483 to 0.5515); and the normalized trace increases by roughly 6.2% (from 0.0582 to 0.0618).

less similar in their attributes to existing occupations than among existing occupations themselves, the measured level of PAM may rise or fall without the structural changes.

To assess the role of structural changes in shaping occupational PAM trends, we construct a counterfactual scenario in which only occupations that are consistently observed across all survey years are retained. Couples in which either spouse holds an occupation not present in every year are excluded from the sample. We then compare the resulting PAM measures to baseline estimates derived from the full sample.

Across all four measures, PAM levels are systematically higher in the restricted sample. For example, under the perfect-random normalization measure, PAM increases from 0.1421 to 0.1744 for the 1955–1959 cohort, and from 0.1610 to 0.1992 for the 1985–1989 cohort.[15]

These results point to two distinct effects. The higher PAM level for the 1985–1989 cohort reflects the impact of excluding newly emerging occupations—indicating that PAM would have been higher had these new occupations not entered the labor market. In contrast, the higher PAM level for the 1955–1959 cohort results from the exclusion of disappearing occupations—suggesting that PAM would also have been higher had those occupations remained. Taken together, these findings suggest that occupational entry and exit over time have attenuated the observed growth in assortative mating, partially obscuring the full extent of rising PAM.

Changes in Occupational Attributes

Occupations are defined by a diverse set of characteristics and skill requirements, many of which evolve over time. These changes can influence the degree of occupational assortative mating. For example, if certain occupations become increasingly differentiated in their reliance on information technology or computer-related skills, their underlying occupational profiles may diverge, reducing measured similarity between spouses in those occupations.

To isolate the impact of changing occupational attributes on trends in assortative mating, we construct a counterfactual scenario in which occupational attribute values are fixed at their

[15] Similar increases are observed for the aggregate likelihood ratio (from 1.0745 to 1.0918 for the 1955–1959 cohort), weighted similarity (from 0.5436 to 0.5635), and normalized trace (from 0.0524 to 0.0732).

baseline levels from 2003. For occupations that appear after 2003, we assign attribute values based on their earliest year of observation. This approach allows us to assess how much of the observed change in PAM is attributable to evolving occupational characteristics, holding all other factors constant.

Under this counterfactual, the perfect-random normalization measure for the 1955–59 cohort declines from 0.1421 to 0.1396, while the measure for the 1985–89 cohort increases from 0.1610 to 0.1623. This implies that changes in occupational attributes over time have dampened the observed rise in PAM by approximately 17 percent.[16]

Taken together, these results indicate that if occupational attributes had remained constant over time, the measured rise in positive assortative mating would have been significantly larger. This suggests that temporal shifts in occupational characteristics have partially obscured the underlying increase in assortative matching.

Rising Average Age of First Marriage

It is well documented that the average age at first marriage in the United States has risen steadily in recent decades. Figure 6 illustrates the proportion of individuals aged 26 to 60 in the CPS who have ever been married, reported by birth cohort. The data show a consistent decline in marriage prevalence over time, with the decline more pronounced among men than among women.

Delayed marriage may influence occupational assortative mating. Individuals who marry later are more likely to have completed their education, established stable careers, and encountered potential spouses in workplace settings, where occupational proximity and lower search costs increase the likelihood of matching with a partner in a similar occupation. Accordingly, the rising age at first marriage may be one contributing factor to the observed increase in occupational PAM across cohorts.

Although the CPS does not report age at first marriage directly, we use age at first birth as a

[16] A similar pattern is observed with the aggregate likelihood ratio: the counterfactual gain (0.0186) exceeds the actual gain (0.0107) by 0.0079, or roughly 42 percent. The weighted similarity and normalized trace measures also show larger counterfactual increases, suggesting that evolving occupational profiles attenuated the observed cohort gains by about 31 percent and 52 percent, respectively.

proxy to examine the relationship between delayed family formation and occupational sorting. We find that women who had their first child at older ages tend to exhibit higher occupational similarity with their husbands. For instance, among women in the 1985–89 cohort, those who had their first child before age 30 have a perfect-random normalization PAM value of 0.1364, whereas those who had their first child after age 30 have a PAM value of 0.1866. This pattern is consistent with the interpretation that delayed childbearing—and by implication, delayed partnership formation—is associated with stronger occupational sorting.

IV Religious Assortative Mating: Second Application

Building on the methodology developed in Section II, this section examines assortative mating in the context of religious affiliation among married couples in the United States. Following the structure used in the occupational PAM analysis (Section III), we describe the relevant data sources, outline the construction of variables, present descriptive statistics, and quantify the extent of religious assortative mating.

4.1. Data Description

Our analysis draws on two primary data sources: the World Values Survey (WVS), which provides comprehensive information on religious beliefs, practices, and values; and the General Social Survey (GSS), which contains detailed data on the religious affiliation of both spouses.

The WVS is a large-scale cross-national survey conducted in seven waves since 1981. It captures a broad range of attitudes and behaviors across social, political, economic, religious, and cultural domains. For this study, we extract 190 religion-related variables that span a wide spectrum of individual beliefs and practices. These variables encompass: frequency of prayer, meditation, and spiritual reflection; belief in God, heaven, hell, and the afterlife; trust in religion to address social, moral, and familial concerns; adherence to religious norms regarding dress and conduct; perspectives on marriage, gender roles, and family structure; and the importance of pilgrimage or sacred rituals.

For couple-level religious affiliation, we utilize the GSS, which uniquely records the religious preferences of both the respondent and their spouse. While the Current Population Survey (CPS) provides a much larger sample, it lacks matched spousal religious data, rendering the GSS the most appropriate dataset for our analysis of religious PAM.

Unlike in the occupational analysis, we do not impose age restrictions on the GSS sample. Religious affiliation tends to remain relatively stable across the life course and is not systematically affected by life-stage transitions such as retirement. As a result, we include all married couples in the GSS who report valid religious affiliation for both spouses.

To align religious groupings across datasets, we harmonize the religious classifications used in the WVS and GSS. The GSS classifies respondents into 13 major religious categories, while the WVS reports religious affiliation using 10 broader categories.[17] Although the most recent WVS wave (2017–2022) includes finer distinctions, we exclude these refinements to ensure consistency across survey waves.

Three of the GSS categories—Inter-nondenominational, Other Eastern Religions, and Native American—are collapsed into a residual "Others" category in the WVS classification. These three groups collectively account for less than 1% of the baseline sample (0.47%, 0.10%, and 0.16%, respectively). Table 6 presents the full distribution of spousal pairings by religious category across all survey years.

While the WVS began surveying U.S. respondents in 1994, the GSS has only consistently reported matched spousal religious affiliations since 2004. To ensure comparability and data consistency, our analysis is restricted to the 2004–2022 period.

4.2. Estimation of Religious Similarity

To reduce the dimensionality of the 190 religion-related variables from the WVS and address potential redundancy, we apply exploratory factor analysis (EFA). This method identifies 25 underlying factors with eigenvalues greater than one, capturing the core dimensions of religious attributes. Together, these factors explain 97.17% of the total variance in the original variables and form the basis of individual-level religious attribute vectors used in our similarity analysis.

As in the occupational PAM analysis, we compute pairwise similarity between religious groups

[17] The ten religious categories in the WVS include: no affiliation, Protestant, Roman Catholic, Orthodox, Jewish, Muslim, Hindu, Buddhist, other Christians, and others.

using cosine similarity applied to the extracted factor vectors. In contrast to the conventional binary approach—which treats only exact matches as indicative of similarity—our similarity-based framework captures varying degrees of alignment between religious traditions. This continuous metric provides a more nuanced and realistic assessment of assortative mating patterns by accounting for partial overlap across belief systems.

A structural distinction arises between the religious and occupational similarity matrices. In the occupational setting, self-similarity values (s_{ii}) are often less than one due to the aggregation of detailed O*NET occupations into broader CPS categories. In the religious context, by contrast, the similarity matrix assigns s_{ii} = 1 for all diagonal elements, reflecting the close correspondence between GSS and WVS religious classifications. Nevertheless, both similarity matrices share key features: they are symmetric and time-varying, adapting to year-specific attribute profiles.

To illustrate the effectiveness of this similarity-based approach, we report the average cosine similarity scores between selected religious categories across all survey years. For example, Catholicism and Protestantism exhibit a high degree of similarity ($s_{C,P}$=0.877), reflecting shared theological foundations and cultural practices. In contrast, Protestantism and Buddhism show lower similarity ($s_{P,B}$=0.495), with Catholicism and Buddhism even lower ($s_{C,B}$=0.493). These gradients of similarity—ignored in binary classification—underscore the value of using a continuous measure. The average pairwise similarity matrix based on cosine similarity is presented in Table 7.

4.3. Measurement of Religious Assortative Mating

Using the matching table constructed in Section 4.1 and the similarity matrix developed in Section 4.2, we apply our framework to examine trends in religious assortative mating (PAM) by birth cohort among married couples in the United States. As in the occupational analysis, we begin with the perfect-random normalization measure, which serves as a benchmark by positioning observed matching patterns between two counterfactual extremes: perfect assortative matching and random matching.

Figures 7 through 10 present cohort-level trends in religious PAM using four measurement approaches: perfect-random normalization, aggregate likelihood ratio, weighted similarity, and normalized trace. All results are reported under cosine similarity. The analysis includes individuals born between 1930 and 1989, grouped into five-year birth cohorts.

Figure 7 shows the results based on the perfect-random normalization measure. Religious PAM declines steadily from older to younger cohorts, with the most pronounced drop occurring between the 1935–1939 and 1940–1944 cohorts. A modest rebound is observed among the most recent cohorts (1980–1984 and 1985–1989).

This decline is consistent with the broader trend of secularization in the United States, particularly since the 1960s. Grant (2008) documents the trajectory of the Aggregate Religiosity Index—a composite measure of religious commitment, belief, and behavior—showing a steady increase until the early 1960s, followed by a sustained decline (see Figure 11).[18] The peak of the religiosity index corresponds closely to the 1935–1939 cohort (in their 20s and 30s between 1959 and 1969), mirroring the peak of religious PAM observed in Figure 7.

We note that the confidence intervals in this figure are relatively wide, reflecting the smaller sample sizes of the WVS and GSS datasets. Nonetheless, the PAM estimates for all cohorts born after 1955 are statistically significantly lower than that for the 1935-1939 birth cohort at the 5% significance level.

In contrast to Figure 7, Figure 8 displays results using the aggregate likelihood ratio measure, which shows a steady increase in religious PAM across cohorts.

Figure 9 displays cohort trends based on the weighted similarity measure. This figure, like Figure 7, shows a general decline in religious PAM across cohorts, though the decline is more moderate. A slight uptick appears among the most recent cohort.

According to equation (9), the divergence in trends between Figures 8 and 9 implies that the weighted similarity under random matching $\sum_{i,j} r_{ij} \cdot s_{ij}$ has declined more rapidly than the observed weighted similarity $\sum_{i,j} a_{ij} \cdot s_{ij}$ for more recent cohorts. This pattern indicates that the marginal distributions of religious affiliation between men and women have become more

[18] The Aggregate Religiosity Index (ARI) is constructed by synthesizing several survey-based indicators of religious behavior and belief, including frequency of religious service attendance, frequency of prayer, self-identified religious affiliation, and the personal importance of religion. Grant (2008) uses a dynamic factor analysis model to combine these indicators into a single latent variable that captures overall levels of aggregate religiosity.

dissimilar over time. Supporting this interpretation, we find that the Euclidean distance between male and female religious distribution vectors increased from 0.04 for the 1930–1934 cohort to 0.07 for the 1985–1989 cohort. The divergence is particularly pronounced in the categories of "no religious affiliation" and "Roman Catholic": the share of men with no affiliation has grown faster than among women, while the share of Catholic men has increased even as it has declined among women.

Although the aggregate likelihood ratio measure (Figure 8) suggests increasing religious PAM—reflecting adjustments for changes in marginal distributions—the perfect-random normalization measure (Figure 7) tells a different story. Once the theoretical limits of perfect and random matching are taken into account, religious PAM appears to have declined. This contrast raises an important question: Which measure provides a more accurate depiction of changes in assortative mating? Given its ability to adjust for both extremes of the matching distribution, the perfect-random normalization measure arguably offers a more robust metric for assessing inter-cohort changes in PAM.

Figure 10 presents the normalized trace measure, which captures only exact religious matches. Its trend closely parallels those in Figures 7 and 9, indicating a persistent decline in exact religious alignment across birth cohorts.[19]

Comparison to the Conventional Approach

Appendix Figures B.1.1 through B.1.3 present the corresponding cohort trends under the conventional identity similarity matrix, aligned with Figures 7, 8, and 10, respectively. Across all four measures, the results under the similarity-based and conventional approaches are remarkably similar. This convergence stems from the structure of the religious similarity matrix: diagonal elements are uniformly equal to one, and the off-diagonal elements are generally low and exhibit limited variation across most religious pairings. Although the off-diagonal values are non-zero, their relatively narrow range limits their influence on similarity-based PAM measures, rendering them functionally close to the identity matrix specification.

[19] Appendix B.2 reports religious PAM results based on Euclidean similarity. Because all diagonal elements equal one in both the cosine and Euclidean similarity matrices, the normalized trace measure yields identical PAM levels under either specification. Moreover, the cohort trends appear nearly identical regardless of the similarity metric used.

V Educational Assortative Mating: Third Application

In this section, we apply our framework to the analysis of educational assortative mating. While education has long been a central focus in the literature on assortative mating (Chiappori et al., 2017; Eika et al., 2019; Gihleb and Lang, 2020; Greenwood et al., 2014; Guell et al., 2015; Shen, 2019), our approach offers a more nuanced perspective. Rather than relying solely on formal educational attainment, we incorporate a broader range of behavioral, health-related, and socioeconomic characteristics associated with educational groups. As in previous sections, the analysis proceeds in three steps: data description, construction of the similarity matrix, and measurement of educational PAM.

5.1. Data Description

Our analysis draws on two primary data sources: the Medical Expenditure Panel Survey (MEPS) and the Current Population Survey (CPS). MEPS provides nationally representative data on health, behavioral, and socioeconomic characteristics, which we use to construct educational group profiles. These profiles are then linked to the CPS, which contains detailed information on spousal education, to analyze patterns of educational assortative mating.

We use MEPS data from 2004 to 2016 and select 34 variables capturing key dimensions of health status, health behaviors, and socioeconomic conditions. These include employment status, personal income, and wages or salary, along with thirty health-related indicators encompassing physical and mental health, health-related beliefs, and smoking behavior—a salient health-risk factor. Together, these variables enable a richer characterization of educational groups, extending beyond formal credentials to include traits likely relevant for partner selection. Appendix Table C.4.1 provides the complete list of variables.

We categorize educational attainment into four groups: less than high school (LHS), high school graduate (HS), some college (SC), and college degree or higher (C). We do not distinguish advanced degrees from bachelor's degrees to ensure consistency across cohorts. While the proportion of individuals with advanced degrees is sufficiently large in older cohorts, educational completion may be right-censored for younger individuals—particularly those born between 1985 and 1989—who had not yet finished schooling at the time of observation. Aggregating college and advanced degrees mitigates this source of bias.

To enhance comparability and interpretability, all variables are rescaled to a common 0–100 scale. Where applicable, coding is adjusted so that higher values represent more favorable outcomes. For example, indicators of poor health (e.g., physical limitations, psychological distress) are reverse-coded, while positive indicators (e.g., excellent self-rated health, no activity limitations) retain their original scale.

Separate attribute vectors are constructed for men and women within each educational group to account for potential gender differences in health, behavior, and socioeconomic profiles—even within the same educational level.[20] As a result, the constructed similarity matrix is asymmetric, unlike the symmetric matrices used in the occupational and religious analyses.

We then apply the similarity matrix to the CPS spousal matching data to measure educational PAM. The CPS sample is restricted to married couples in which both spouses are aged 25 or older, consistent with the assumption that most individuals have completed formal education by that age.[21] This age restriction differs from the occupational PAM analysis, where we limited the sample to those aged 26–60 to reflect active labor force participation. Since education is time-invariant once completed, we impose no upper age bound here.

Table 8 presents the aggregate matching table of spousal educational pairings from the CPS across the full sample period. While the table shows a strong concentration along the diagonal—indicating widespread educational homogamy—it also reveals that more than 1.3 million of 4.7 million couples consist of partners with different education levels. Our framework gives analytical weight to these off-diagonal pairings, treating them not as residual noise but as informative variation that can shed light on the broader structure of educational assortative mating.

---

[20] Such differentiation was not feasible in the occupational analysis, due to the absence of gender-specific information in O*NET, nor in the religious analysis, where limited sample sizes in the WVS prevented reliable estimation of gender-disaggregated religious attribute profiles.

[21] The age threshold of 25 is a standard benchmark in educational statistics and is widely used by institutions such as the U.S. Census Bureau, the National Center for Education Statistics (NCES), and the OECD.

5.2. Estimation of Educational Similarity

The educational similarity matrix is constructed using the 34 attributes derived from the MEPS dataset, as outlined in Section 5.1. These attributes span key domains—including health status, health behaviors, and socioeconomic outcomes—that collectively characterize individuals across educational attainment levels.

To reduce dimensionality and enhance interpretability, we apply exploratory factor analysis to the full set of attributes and extract five latent factors, retaining only those with eigenvalues greater than one. These factors summarize the underlying variation across educational groups and jointly explain 90.63% of the total variance in the original variables. The resulting factor scores form the basis for constructing group-level educational profiles used in the similarity analysis.

Using these scores, we construct separate attribute vectors for men and women within each of the four educational groups: less than high school, high school graduate, some college, and college degree or higher. This gender-specific construction reflects the empirical reality that individuals with the same level of formal education may differ meaningfully in behavioral, health, and economic characteristics depending on gender. Prior research has consistently documented such differences in labor market outcomes, health trajectories, and lifestyle behaviors (Bianchi et al., 2006; Montez et al., 2019; Schnittker, 2004).

Following the approach used in earlier applications, we compute cosine similarity between these gender-specific attribute vectors to construct a similarity matrix for each survey year. In contrast to the occupational and religious analyses, which rely on repeated cross-sectional data, the educational analysis is based on panel data. Consequently, the time-varying similarity matrices for education may capture not only secular changes in educational profiles but also within-cohort, age-related variation.[22]

[22] The within-cohort, age-related variation is an effect which we do not want to include in the analysis. To address this concern, we conducted a robustness check by constructing a single, time-invariant similarity matrix using pooled observations from all survey years. This aggregate similarity matrix captures the overall structure of educational similarity while minimizing confounding age-related variation. The resulting cohort patterns of educational PAM, based on this time-invariant similarity matrix, are qualitatively similar to those obtained using year-specific matrices (see Appendix C.3).

The resulting similarity matrices are asymmetric and do not feature uniform diagonal elements. That is, the similarity between men with education level i and women with level j may differ from the reverse pairing, and the self-similarity term s_{ii} is not fixed at one. This construction contrasts with the religious similarity matrix, which is symmetric with s_{ii} = 1 by design, and the occupational similarity matrix, which is also symmetric but may include s_{ii} < 1 due to aggregation across heterogeneous occupational categories.

Table 9 presents the average similarity matrix over the sample period based on cosine similarity. The matrix displays strong diagonal dominance, indicating high similarity among educationally homogamous couples. More importantly, the off-diagonal elements vary systematically, revealing that educational mismatches differ in their degrees of dissimilarity and the similarity scores monotonically decline as the educational gap between spouses widens. For instance, a pairing between a high school–educated husband and a wife who did not complete high school yields a relatively high similarity score (0.864), whereas a match between a college-educated husband and a wife with less than high school education records the lowest similarity in the matrix (0.029).

These patterns reinforce a central insight of our framework: assortative mating should not be evaluated solely based on exact matches. Rather, meaningful variation exists across all possible pairings, and accounting for this continuous structure allows for a richer and more accurate assessment of matching behavior than binary categorizations can provide.

5.3. Measurement of Educational Assortative Mating

Our baseline sample includes married couples in which both spouses report valid educational attainment. Cohort classification is based on the husband's year of birth, which ranges from 1919 to 1991, corresponding to five-year birth cohorts spanning from 1915–1919 to 1990–1994. However, the earliest (1915–1919 and 1920–1924) and the latest (1990–1994) cohorts each constitute less than 1% of the sample (0.06%, 0.74%, and 0.09%, respectively), rendering estimates for these groups unreliable. We therefore restrict our analysis to husbands born between 1925 and 1989 to ensure broader coverage and more robust cohort-level estimates.

Figures 12 through 15 present cohort-level trajectories of educational PAM, evaluated using the four similarity-based measures introduced earlier: perfect-random normalization, aggregate likelihood ratio, weighted similarity, and normalized trace. For each measure, results are reported based on cosine similarity. The analysis uses CPS spousal pairings, with educational PAM computed separately for each five-year birth cohort.

Figure 12 shows the trajectory of educational PAM using the perfect-random normalization measure. PAM remains relatively stable for earlier cohorts (born in the late 1920s through early 1950s), followed by a notable decline for the late 1950s cohort, and then a marked upward trend for cohorts born after 1960. Because the perfect-random normalization measure controls for changes in the marginal distributions of men's and women's educational attainment, it isolates changes in the underlying preference for educational similarity. The results suggest that this preference has strengthened among more recent cohorts—a pattern consistent with Chiappori et al. (2025), who document a resurgence in educational assortative mating among the 1970s cohort in comparison to the 1950s cohort.

Figure 13 presents results based on the aggregate likelihood ratio measure. The cohort pattern here broadly mirrors that in Figure 12: PAM remains stable before the late 1960s cohort and rises thereafter. However, the magnitude of the increase is smaller, indicating that the aggregate likelihood ratio measure provides a more conservative estimate of inter-cohort variation in educational PAM.

As in the religious domain, we observe a clear divergence between the perfect-random normalization and aggregate likelihood ratio measures. While the former indicates substantial changes in educational sorting across cohorts, the latter points to more modest shifts. This contrast underscores the importance of measurement choice in interpreting long-term trends in assortative mating.

Figure 14 presents cohort-level trends in educational PAM based on the weighted similarity measure, again using cosine similarity. The results show a gradual increase from the 1925–1929 to the 1985–1989 cohorts. Between the 1955–1959 and 1985–1989 cohorts, the weighted similarity measure rises by 8.5%, whereas the perfect-random normalization measure records a larger increase of 24.8% over the same interval.

Figure 15 displays results from the normalized trace measure, which captures the proportion of couples with identical education levels. This measure shows a steady, monotonic increase in

PAM across cohorts, with no sign of reversal or plateauing.

Together, the results in Figures 14 and 15 highlight a key insight: failing to account for changes in the marginal distributions of educational attainment between men and women can yield trend patterns that differ substantially from those produced by methods that explicitly incorporate counterfactual benchmarks such as perfect and random matching. In particular, the narrowing gender gap in educational attainment over the past century has mechanically increased the frequency of college-college pairings, producing a monotonic upward trend in educational PAM as captured by indices like weighted similarity and normalized trace.

Comparison to the Conventional Approach

To further contextualize our findings, Appendix Figures C.1.1 through C.1.3 present benchmark trends based on conventional measures of educational PAM that assume an identity similarity matrix. Appendix Figure C.1.1, which displays the conventional perfect-random normalization measure, shows a rising trend in PAM after the 1955-1959 cohort, closely mirroring the pattern in Figure 12. However, unlike Figure 12, it also shows a decline in PAM for earlier cohorts.

Appendix Figure C.1.2 reports the conventional aggregate likelihood ratio measure. The overall trajectory resembles that of Figure 13, though it indicates somewhat higher levels of PAM among cohorts born before 1950.

Appendix Figure C.1.3 presents results from the normalized trace measure under the conventional framework. It shows a steadily increasing trend in educational PAM across all cohorts, with a pattern and magnitude of change that closely align with those reported in Figure 15.

VI. Concluding Remarks

This paper introduces a new framework for measuring positive assortative mating (PAM) that moves beyond the conventional binary classification of partner matches. Rather than treating couples as either perfectly sorted or entirely dissimilar, we develop a continuous similarity-based approach that uses cosine similarity to capture the degree of alignment across multidimensional trait compositions. Within this framework, we construct similarity-weighted matching matrices and adapt widely used PAM indices—including the normalized trace, aggregate likelihood ratio, and perfect-random normalization—to account for partial similarity across types. For comparison, we also report results from conventional identity matrix-based measures.

We apply this framework to three key traits in partner selection—occupation, religion, and education—and uncover patterns of assortative mating that conventional binary measures fail to capture. Our analysis shows that identity-based metrics, which assume complete similarity only for exact trait matches and zero otherwise, often misrepresent both the level and trajectory of PAM, particularly when traits are multi-categorical or exhibit continuous variation.

In the occupational domain, similarity-based PAM measures exhibit a pronounced U-shaped pattern across birth cohorts, in contrast to the flat or muted trends produced by traditional metrics. This highlights the value of incorporating similarity information into PAM measurement.

In the religious domain, we observe a general decline in assortative mating across birth cohorts, reflecting the broader secularization trends in American society. However, the aggregate likelihood ratio measure shows a countervailing upward trend, underscoring the sensitivity of PAM trends to the choice of metric—particularly in settings where group distributions evolve over time.

In the educational domain, we find a marked rise in PAM among cohorts born after the 1960s, reflecting both a growing preference for educational similarity and the convergence in male and female educational attainment.

Taken together, our findings demonstrate that the similarity-based framework offers a flexible, theoretically grounded, and empirically robust method for analyzing assortative mating. By allowing for partial alignment across trait types, this approach improves upon conventional practices and offers new insights into the formation of families, the dynamics of social stratification, and the transmission of advantage across generations. Future extensions may apply this framework to other partner traits, examine non-marital unions, or explore international variation in assortative mating processes.

This study also raises important limitations that we aim to address in future work. One limitation concerns the potential arbitrariness in selecting the attributes used to define similarity. Although we rely on high-quality datasets—O*NET for occupation, the WVS for religion, and MEPS for education—and reduce dimensionality using exploratory factor analysis, the resulting similarity matrices still depend on researcher judgment. Different choices of attributes or factor structures may yield different similarity scores, potentially altering the observed trends in PAM.

Several strategies could help address this concern. First, sensitivity analyses could test how results vary with alternative attribute sets or dimensionality reduction techniques, such as principal component analysis or supervised learning methods. Second, one could estimate similarity weights directly from observed matches using structural models of partner choice, generating an endogenous measure of trait alignment. Such approaches would bolster the robustness and interpretability of similarity-based PAM estimates.

A second limitation relates to the endogeneity of trait formation. For both education and occupation, partner traits may evolve within the marriage itself. Educational attainment may be completed after marriage, and occupational choices may shift due to joint decisions—such as relocation or childrearing. These dynamics make it difficult to isolate assortative matching at the point of mate selection from post-matching convergence. Accordingly, our measures reflect observed similarity at the time of survey, which may differ from initial matching conditions.

Despite these limitations, our similarity-based framework offers a valuable tool for studying assortative mating in a richer and more flexible way. We hope it will encourage further empirical applications and theoretical developments in the study of household formation and inequality.

References

Acemoglu, Daron, Nicolaj Søndergaard Mühlbach, and Scott Andrew J. 2022. "The Rise of Age-Friendly Jobs." *The Journal of the Economics of Ageing*, 23.

Becker, Gary S. 1981. *A Treatise on the Family*. Cambridge, MA: Harvard University Press.

Bianchi, Suzanne M., John P. Robinson, and Melissa A. Milkie. 2006. *Changing Rhythms of American Family Life*. New York: Russell Sage Foundation.

Blau, Francine D., Marianne A. Ferber, and Anne E. Winkler. 2006. *The Economics of Women, Men, and Work*. 5th ed. Upper Saddle River, NJ: Prentice Hall.

Bureau of Labor Statistics. 2017. "A Look at Women's Education and Earnings Since the 1970s." *The Economics Daily*, U.S. Department of Labor. https://www.bls.gov/opub/ted/2017/a-look-at-womens-education-and-earnings-since-the-1970s.htm (accessed May 14, 2025).

Chiappori, Pierre-André, Monica Costa Dias, and Costas Meghir. 2020. "Changes in Assortative Matching: Theory and Evidence for the US." *NBER Working Paper* No. 26932.

Chiappori, Pierre-André, Monica Costa Dias, Costas Meghir, and Hanzhe Zhang. 2025. "Changes in Marital Sorting: Theory and Evidence from the US." *Journal of Political Economy*, forthcoming.

Chiappori, Pierre-André, Bernard Salanié, and Yoram Weiss. 2017. "Partner Choice, Investment in Children, and the Marital College Premium." *American Economic Review*, 107 (8): 2109–2167.

Dahlke, Jeffrey A., David J. Putka, Olivia R. Shewach, and Paul Lewis. 2022. "Developing Related Occupations for the O*NET Program." HumRRO Report No. 036. https://www.onetcenter.org/reports/Related_2022.html.

Eika, Lasse, Magne Mogstad, and Basit Zafar. 2019. "Educational Assortative Mating and Household Income Inequality." *Journal of Political Economy*, 127 (6): 2795–2835.

Grant, J. Tobin. 2008. "Measuring Aggregate Religiosity in the United States, 1952–2005." *Sociological Spectrum*, 28 (5): 460–476.

Greenwood, Jeremy, Nezih Guner, Georgi Kocharkov, and Cezar Santos. 2014. "Marry Your Like: Assortative Mating and Income Inequality." *American Economic Review: Papers and Proceedings*, 104 (5): 348–353.

Güell, Maia, José V. Rodríguez Mora, and Christopher I. Telmer. 2015. "The Informational Content of Surnames, the Evolution of Intergenerational Mobility, and Assortative Mating." *Review of Economic Studies*, 82 (2): 693–735.

Han, Shiyu, and Yingyi Qian. 2021. "Concentration and Dispersion: School-to-Work Linkages and Their Impact on Occupational Assortative Mating." *The Social Science Journal*, 1–18.

Heckman, James J., Rodrigo Pinto, and Peter Savelyev. 2013. "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review*, 103 (6): 2052–2086.

Henning, Martin, Rikard Eriksson, Petter Garefelt, Henrik Martin, and Zoltán Elekes. 2025. "Job Relatedness, Local Skill Coherence and Economic Performance: A Job Postings Approach." *Regional Studies, Regional Science*, 12 (1): 95–122.

Hyatt, Henry R. 2019. "Coworking Couples and the Similar Jobs of Dual-Earner Households." *Monthly Labor Review*, 142: 1.

Jacob, Nicolas. 2024. "Assortative Mating, Marital Stability and the Role of Business Cycles in the United States from 1968 to 2011." *Journal of Applied Economics*, 27 (1).

Makridis, Christos, Luke Hickman, and Brian Manning. 2023. "Earnings Are Greater and Increasing in Occupations That Require Intellectual Tenacity." *Social Science Research Network*. https://doi.org/10.2139/ssrn.XXXXX.

Mansour, Hani, and Terra McKinnish. 2018. "Same-Occupation Spouses: Preferences or Search Costs?" *Journal of Population Economics*, 31: 1005–1033.

Montez, Jennifer Karas, Mark D. Hayward, and Anna Zajacova. 2019. "Educational Attainment and Adult Mortality in the United States: A Systematic Analysis of Functional Form." *Demography*, 56 (6): 2053–2080.

Okumura, Tsunao, and Emiko Usui. 2016. "Intergenerational Transmission of Skills and Differences in Labor Market Outcomes for Blacks and Whites." In *Inequality: Causes and Consequences*, 227–286. Emerald Group Publishing.

Robinson, Cervin. 2018. "Occupational Mobility, Occupation Distance, and Specific Human Capital." *Journal of Human Resources*, 53 (2): 513–551.

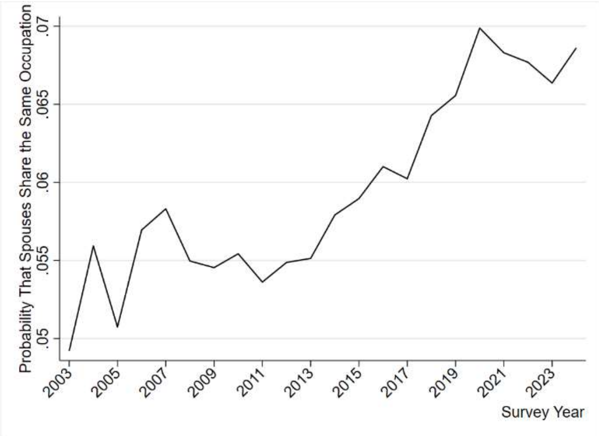
Schnittker, Jason. 2004. "Education and the Changing Shape of the Income Gradient in Health." *Journal of Health and Social Behavior*, 45 (3): 286–305.

Schwartz, Christine R. 2013. "Trends and Variation in Assortative Mating: Causes and Consequences." *Annual Review of Sociology*, 39 (1): 451–470.

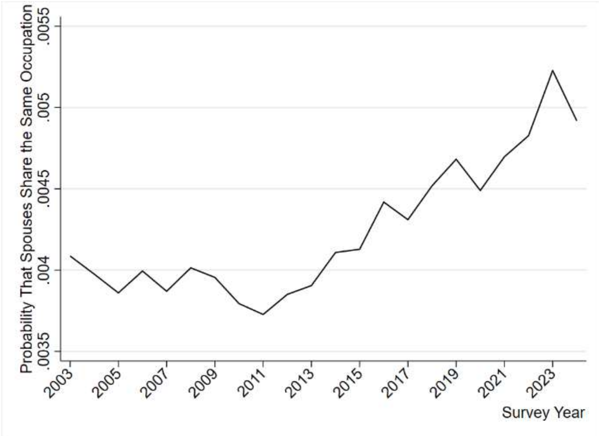
Schwartz, Christine R., Yujin Wang, and Robert D. Mare. 2021. "Opportunity and Change in Occupational Assortative Mating." *Social Science Research*, 99: 102600.

Van Kammen, Barbara, and Samuel J. Adams. 2014. "Dissimilar Occupations and Marital Stability." *IZA Journal of Labor Economics*, 3: 1–18.

Figure 1. Trends in Spousal Occupational Matching Probability



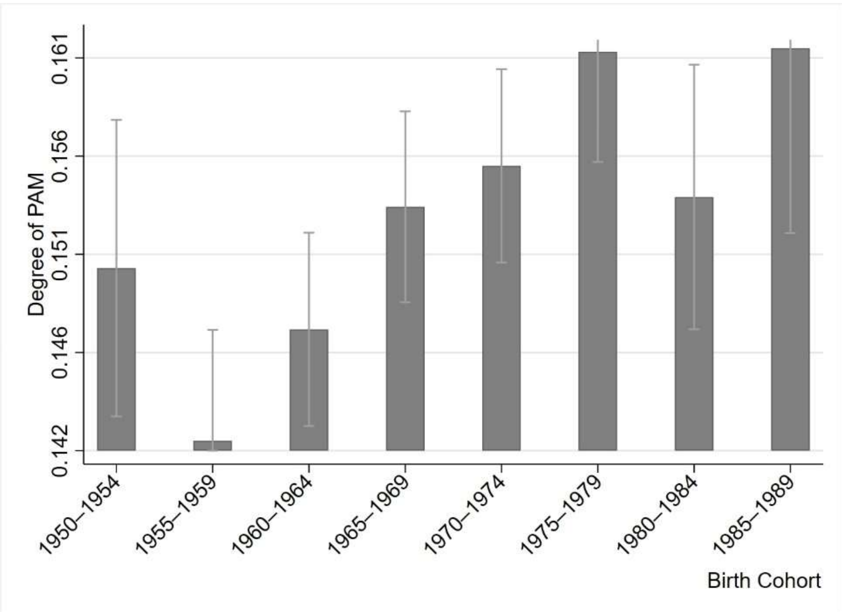
A. Observed Matches



B. Random Matches

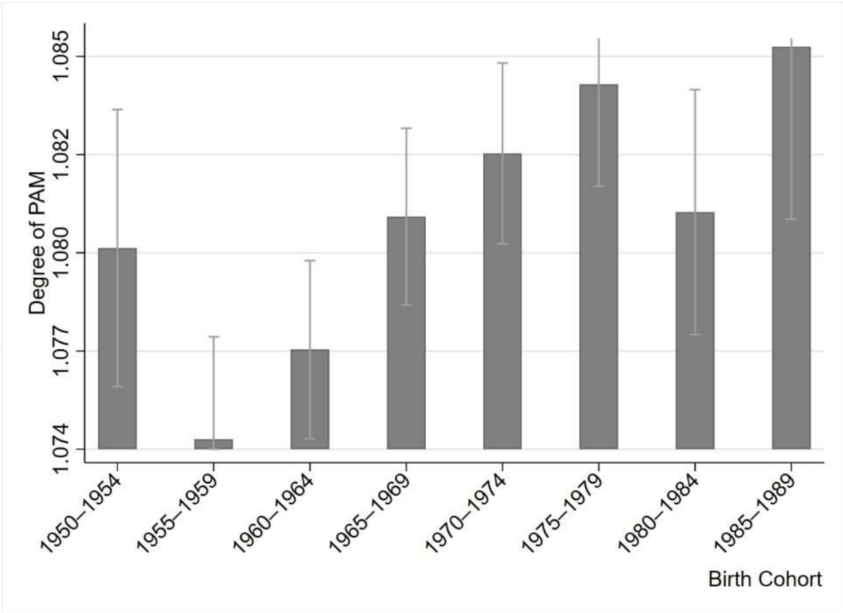
Note: Panel A plots the observed probability that a married couple shares the same occupation by survey year. Panel B presents the corresponding probability under random matching, holding marginal occupational distributions constant. The analysis is based on CPS data from 2003 to 2024 and includes only married couples.

Figure 2. Trends in Occupational PAM across Birth Cohorts Using the Perfect-Random Normalization Measure



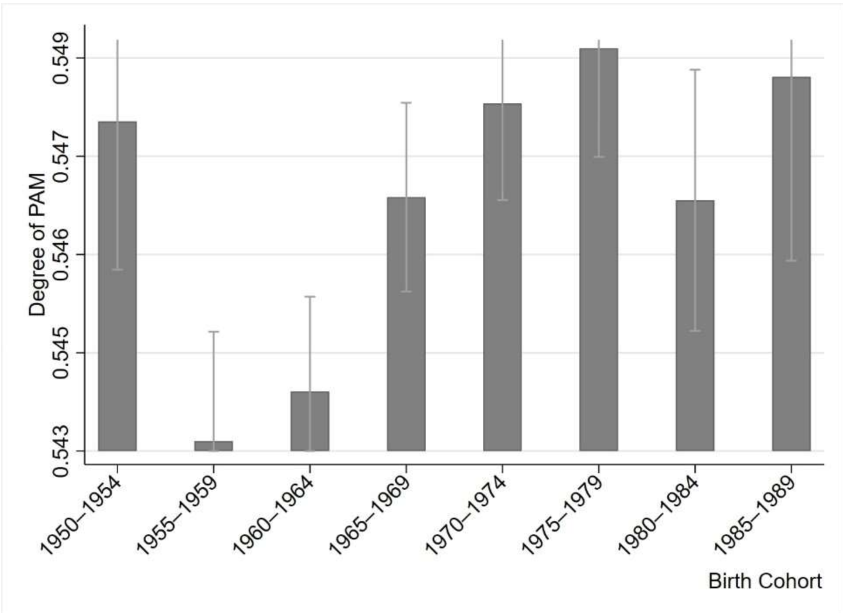
Note: This figure presents occupational PAM across birth cohorts, measured using the perfect-random normalization approach. The similarity matrix is constructed from O*NET occupational attributes using cosine similarity and linked to spouse occupations reported in the CPS data. Vertical lines present 95% confidence intervals based on standard errors estimated via the delta method. Confidence intervals whose upper or lower bounds exceed the plotted range are truncated.

Figure 3. Trends in Occupational PAM across Birth Cohorts
Using the Aggregate Likelihood Ratio Measure



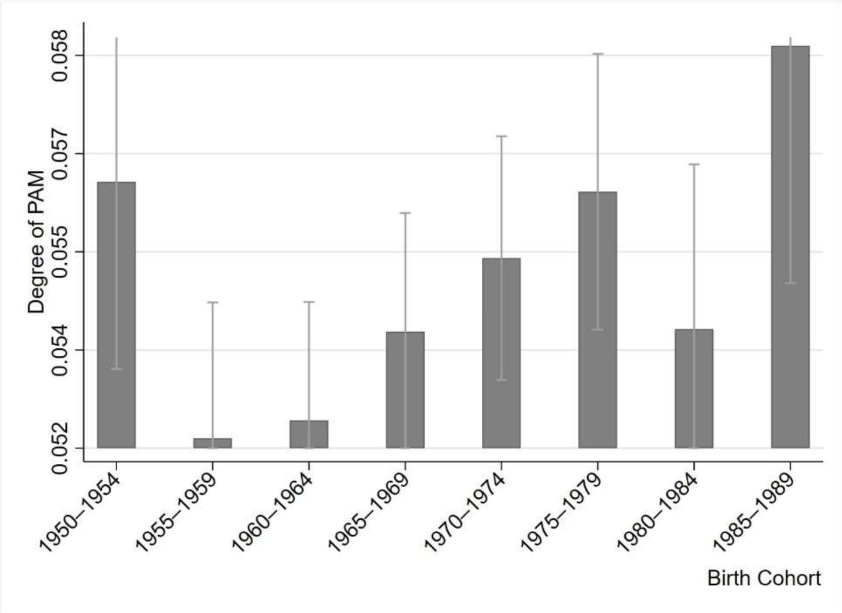
Note: This figure presents occupational PAM across birth cohorts, measured using the aggregate likelihood ratio approach. The similarity matrix is constructed from O*NET occupational attributes using cosine similarity and is linked to spouse occupations reported in the CPS data. Vertical lines present 95% confidence intervals based on standard errors estimated via the delta method. Confidence intervals whose upper or lower bounds exceed the plotted range are truncated.

Figure 4. Trends in Occupational PAM across Birth Cohorts
Using the Weighted Similarity Measure



Note: This figure presents occupational PAM across birth cohorts, measured using the weighted similarity approach. The similarity matrix is constructed from O*NET occupational attributes using cosine similarity and is linked to spouse occupations reported in the CPS data. Vertical lines present 95% confidence intervals based on standard errors estimated via the delta method. Confidence intervals whose upper or lower bounds exceed the plotted range are truncated.

Figure 5. Trends in Occupational PAM across Birth Cohorts Using the Normalized Trace Measure

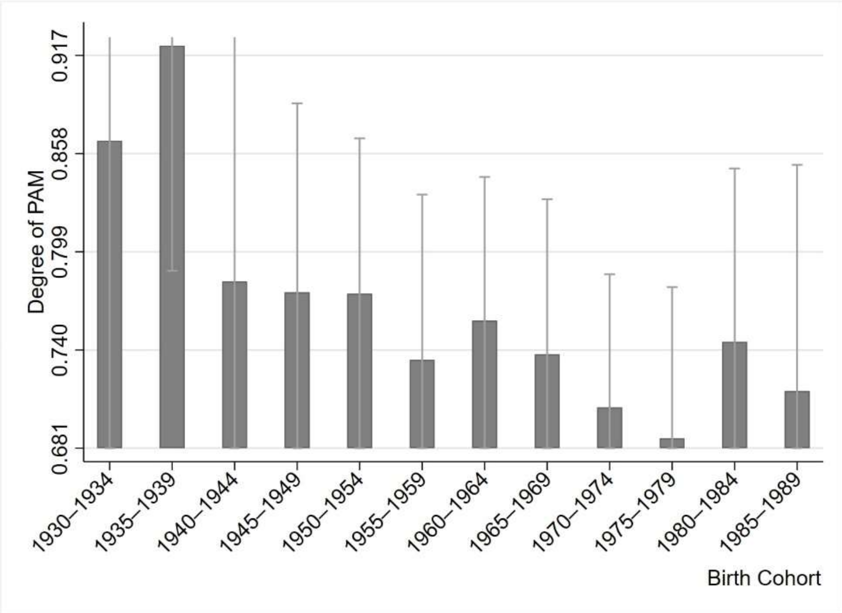


Note: This figure presents occupational PAM across birth cohorts, measured using the normalized trace approach. The similarity matrix is constructed from O*NET occupational attributes using cosine similarity and is linked to spouse occupations reported in the CPS data. Vertical lines present 95% confidence intervals based on standard errors estimated via the delta method. Confidence intervals whose upper or lower bounds exceed the plotted range are truncated.

Figure 6. Trends in Ever-Married Prevalence, CPS (1950–1989 Birth Cohorts)

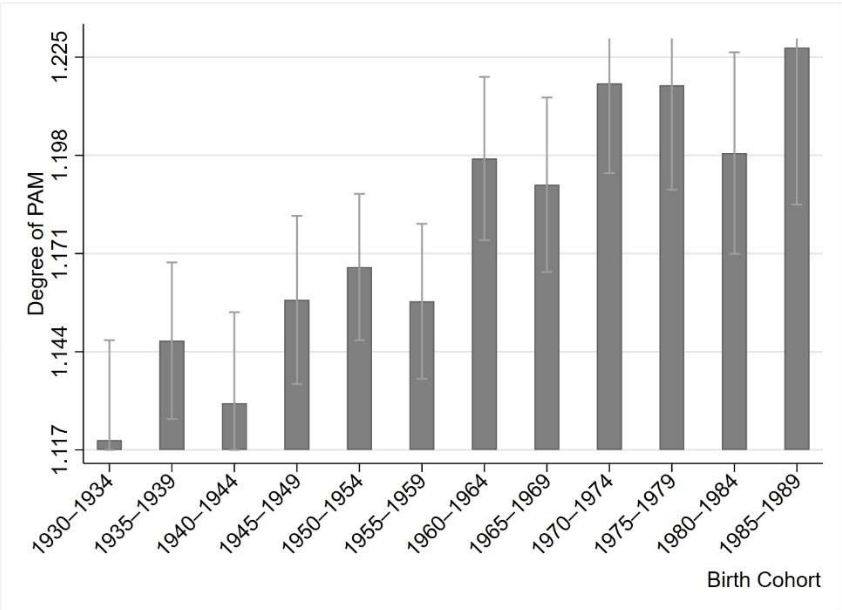
Note: This figure shows, for each five-year birth cohort from 1950–54 to 1985–89, the share of CPS respondents aged 26–60 who have ever married, by sex. The solid line denotes women and the dashed line denotes men.

Figure 7. Trends in Religious PAM across Birth Cohorts
Using the Perfect-Random Normalization Measure



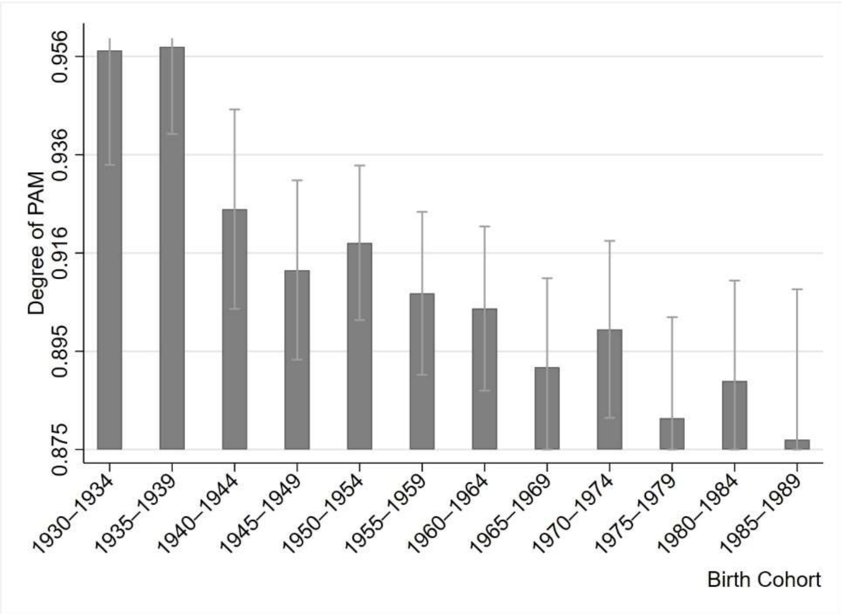
Note: This figure presents religious PAM across birth cohorts using the perfect-random normalization approach. Cosine similarity scores are derived from religious attribute vectors constructed from the WVS and applied to married couples observed in the GSS. Vertical lines present 95% confidence intervals based on standard errors estimated via the delta method. Confidence intervals whose upper or lower bounds exceed the plotted range are truncated.

Figure 8. Trends in Religious PAM across Birth Cohorts
Using the Aggregate Likelihood Ratio Measure



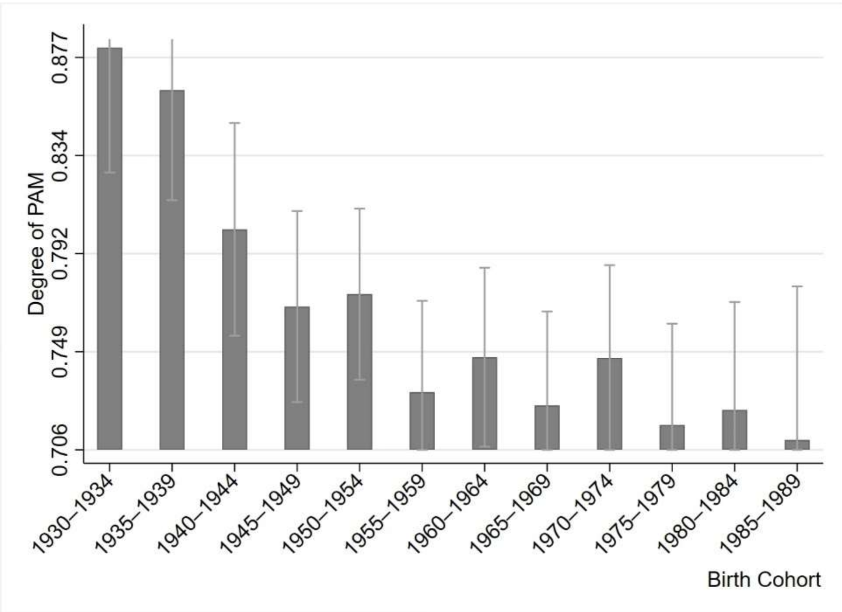
Note: This figure presents religious PAM across birth cohorts using the aggregate likelihood ratio approach. Cosine similarity scores are derived from religious attribute vectors constructed from the WVS and applied to married couples observed in the GSS. Vertical lines present 95% confidence intervals based on standard errors estimated via the delta method. Confidence intervals whose upper or lower bounds exceed the plotted range are truncated.

Figure 9. Trends in Religious PAM across Birth Cohorts
Using the Weighted Similarity Measure



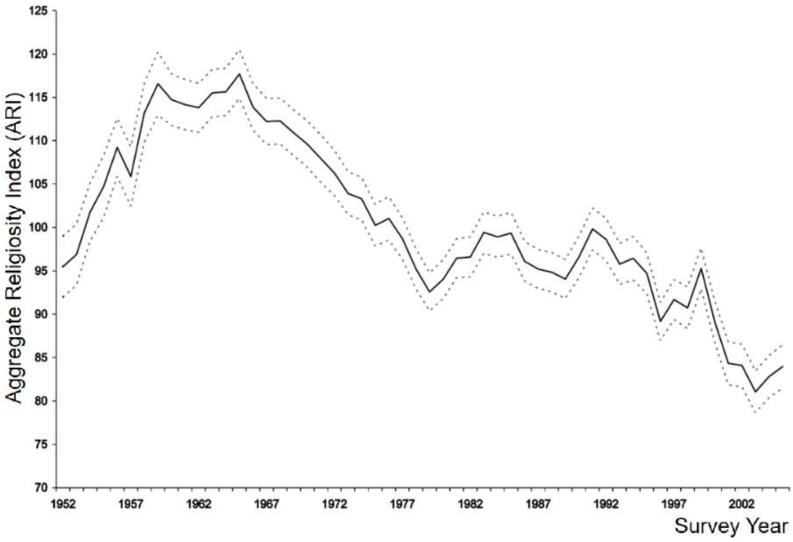
Note: This figure presents religious PAM across birth cohorts using the weighted similarity approach. Cosine similarity scores are derived from religious attribute vectors constructed from the WVS and applied to married couples observed in the GSS. Vertical lines present 95% confidence intervals based on standard errors estimated via the delta method. Confidence intervals whose upper or lower bounds exceed the plotted range are truncated.

Figure 10. Trends in Religious PAM across Birth Cohorts
Using the Normalized Trace Measure



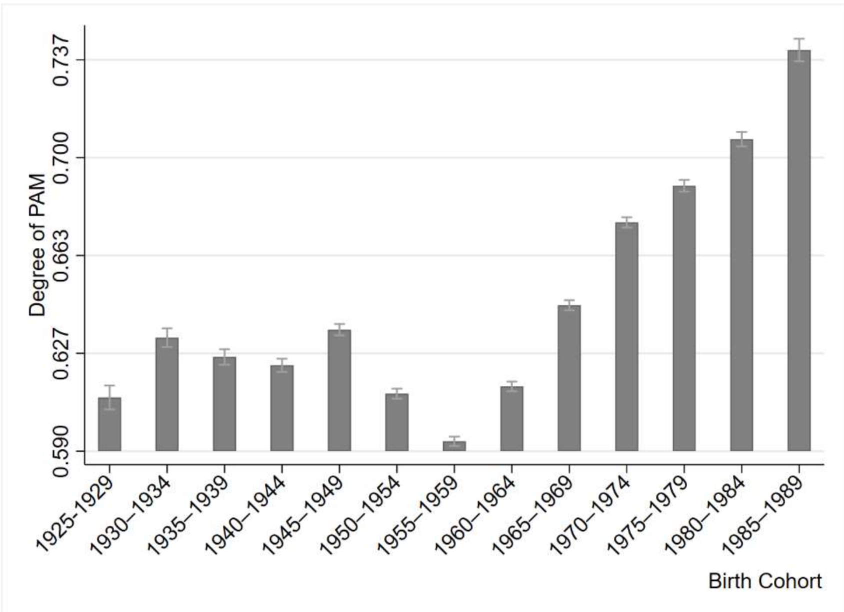
Note: This figure presents religious PAM across birth cohorts using the normalized trace approach. Cosine similarity scores are derived from religious attribute vectors constructed from the WVS and applied to married couples observed in the GSS. Vertical lines present 95% confidence intervals based on standard errors estimated via the delta method. Confidence intervals whose upper or lower bounds exceed the plotted range are truncated.

Figure 11. Aggregate Religiosity Index (ARI), 1952–2003



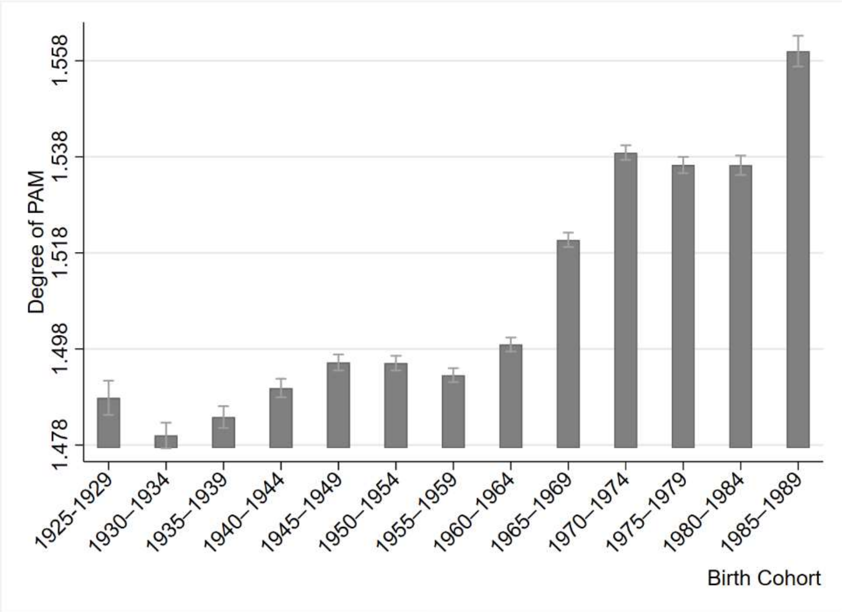
Note: The solid line shows the estimated Aggregate Religiosity Index (ARI) from a time-varying dynamic factor model that combines four survey-based indicators: frequency of religious service attendance, frequency of prayer, self-identified religious affiliation, and personal importance of religion. Dotted lines indicate ±1 standard error bands. The index is normalized to 100 in 1952 for comparability. Source: Grant, J. T. (2008).

Figure 12. Trends in Educational PAM across Birth Cohorts
Using the Perfect-Random Normalization Measure



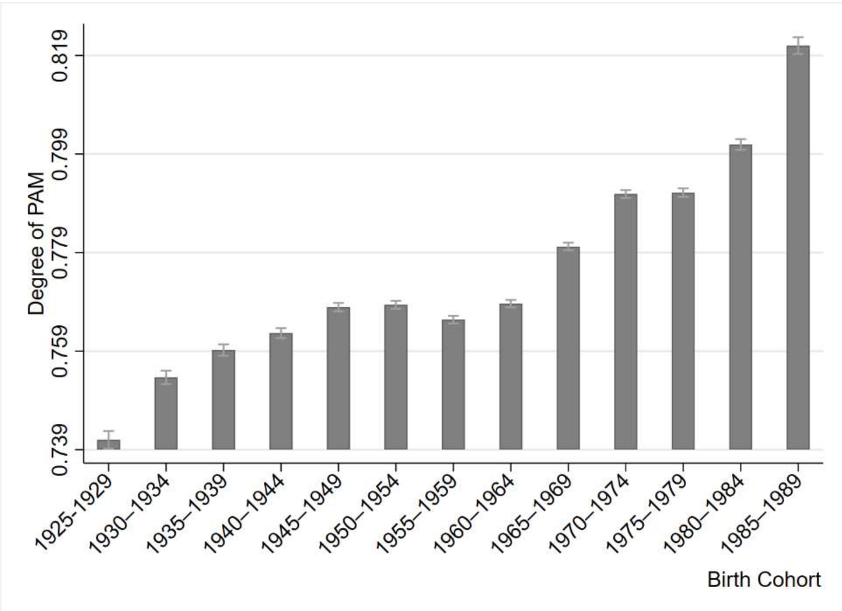
Note: This figure presents educational PAM across birth cohorts, measured using the perfect-random normalization approach. Cosine similarity scores are derived from attribute vectors constructed from the MEPS and applied to married couples observed in the CPS. Vertical lines present 95% confidence intervals based on standard errors estimated via the delta method.

Figure 13. Trends in Educational PAM across Birth Cohorts
Using the Aggregate Likelihood Ratio Measure



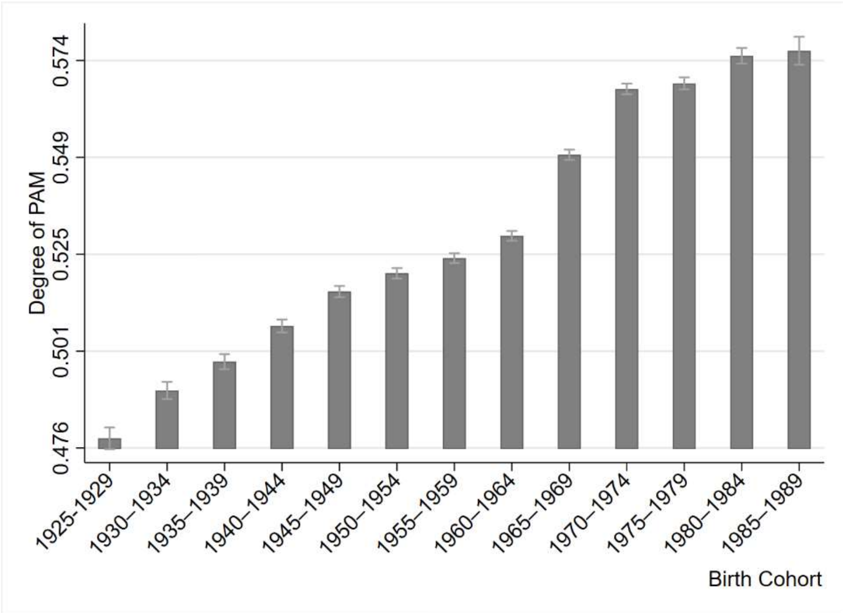
Note: This figure presents educational PAM across birth cohorts, measured using the aggregate likelihood ratio approach. Cosine similarity scores are derived from attribute vectors constructed from the MEPS and applied to married couples observed in the CPS. Vertical lines present 95% confidence intervals based on standard errors estimated via the delta method.

Figure 14. Trends in Educational PAM across Birth Cohorts
Using the Weighted Similarity Measure



Note: This figure presents educational PAM across birth cohorts, measured using the weighted similarity approach. Cosine similarity scores are derived from attribute vectors constructed from the MEPS and applied to married couples observed in the CPS. Vertical lines present 95% confidence intervals based on standard errors estimated via the delta method.

Figure 15. Trends in Educational PAM across Birth Cohorts
Using the Normalized Trace Measure



Note: This figure presents educational PAM across birth cohorts, measured using the normalized trace approach. Cosine similarity scores are derived from attribute vectors constructed from the MEPS and applied to married couples observed in the CPS. Vertical lines present 95% confidence intervals based on standard errors estimated via the delta method.

Table 4. Distribution of Pairwise Occupational Cosine Similarity Scores

Range	Share (%)
s = 1	0.09
0.9 ≤ s < 1	0.08
0.8 ≤ s < 0.9	0.78
0.7 ≤ s < 0.8	3.61
0.6 ≤ s < 0.7	10.79
0.5 ≤ s < 0.6	29.70
0.4 ≤ s < 0.5	41.58
0.3 ≤ s < 0.4	12.55
0.2 ≤ s < 0.3	0.80
0.1 ≤ s < 0.2	0.01

Mean Similarity: 0.502 **Standard Deviation:** 0.102

Table 5. Illustrative Subset of Occupational Cosine Similarity Matrix

Occupation	Physician	Dentist	Math	Econ	RecWork	Trucker	Parking	TourGuide
Physician	0.845	0.792	0.506	0.492	0.487	0.432	0.397	0.375
Dentist	0.792	0.921	0.422	0.430	0.395	0.444	0.286	0.384
Math	0.506	0.422	0.984	0.842	0.351	0.497	0.416	0.458
Econ	0.492	0.430	0.842	0.988	0.286	0.540	0.469	0.478
RecWork	0.487	0.395	0.351	0.286	1.000	0.399	0.536	0.660
Trucker	0.432	0.444	0.497	0.540	0.399	0.894	0.752	0.629
Parking	0.397	0.286	0.416	0.469	0.536	0.752	1.000	0.686
TourGuide	0.375	0.384	0.458	0.478	0.660	0.629	0.686	0.863

Note: This table reports cosine similarity values among a selected subset of occupations, abbreviated as follows: Physician (Physicians and Surgeons), Dentist, Math (Mathematicians), Econ (Economists), RecWork (Recreation Workers), Trucker (Truck Drivers), Parking (Parking Enforcement Workers), and TourGuide (Tour and Travel Guides). Similarities are drawn from the full occupational similarity matrix, constructed using standardized factor scores based on 148 worker-oriented attributes from the O*NET database. The full matrix is available upon request.

Table 6. Spousal Religious Pairings in the GSS Sample: 2004–2022 (Aggregate Observations)

	None	Catholic	Protestant	Orthodox	Jew	Muslim	Hindu	Buddhist	Other Christian	Other
None	1,277	365	570	5	37	5	6	34	39	49
Catholic	165	2,227	476	12	16	2	1	9	29	14
Protestant	231	446	4,982	5	21	4	2	10	70	44
Orthodox	6	9	7	26	0	0	0	0	0	10
Jew	18	23	30	1	151	0	0	0	5	2
Muslim	1	3	13	1	0	62	2	0	0	4
Hindu	5	4	4	1	1	1	77	0	0	0
Buddhist	16	3	8	0	1	0	0	36	2	3
Other Christian	19	20	72	1	0	0	0	0	189	4
Other	29	21	40	5	1	2	1	5	6	82

Note: This table reports the distribution of religious pairings among married couples in the GSS sample from 2004 to 2022. Each cell shows the total number of couples in which the husband and wife belong to the corresponding religious categories, aggregated across all survey waves. Rows indicate the husband's religion; columns indicate the wife's.

Table 7. Average Religious Cosine Similarity Scores Across the Full Sample, 2004–2022

	None	Catholic	Protestant	Orthodox	Jew	Muslim	Hindu	Buddhist	Other Christian	Other
None	1.000	0.638	0.451	0.511	0.600	0.523	0.511	0.511	0.497	0.628
Catholic	0.638	1.000	0.877	0.517	0.584	0.532	0.525	0.493	0.479	0.586
Protestant	0.451	0.877	1.000	0.485	0.545	0.533	0.489	0.496	0.477	0.624
Orthodox	0.511	0.517	0.485	1.000	0.454	0.498	0.515	0.476	0.456	0.559
Jew	0.600	0.584	0.545	0.454	1.000	0.500	0.535	0.512	0.433	0.466
Muslim	0.523	0.532	0.533	0.498	0.500	1.000	0.473	0.486	0.478	0.500
Hindu	0.511	0.525	0.489	0.515	0.535	0.473	1.000	0.498	0.466	0.518
Buddhist	0.511	0.493	0.496	0.476	0.512	0.486	0.498	1.000	0.506	0.482
Other Christian	0.497	0.479	0.477	0.456	0.433	0.478	0.466	0.506	1.000	0.443
Other	0.628	0.586	0.624	0.559	0.466	0.500	0.518	0.482	0.443	1.000

Note: This table reports average pairwise cosine similarity scores between religious categories, calculated across the full GSS sample from 2004 to 2022. Each cell represents the mean similarity between the corresponding religious affiliations.

Table 8. Spousal Educational Pairings in the CPS Sample, 2004–2016: Aggregate Observations

Educational attainment	LHS	HS	SC	C	Total
LHS	217,208	156,840	57,052	17,455	448,555
HS	111,310	776,696	356,311	185,820	1,430,137
SC	33,807	316,698	525,761	295,307	1,171,573
C	12,406	182,817	345,442	1,083,299	1,623,964
Total	374,731	1,433,051	1,284,566	1,581,881	4,674,229

Note: This table reports the distribution of educational pairings among married couples in the CPS data from 2004 to 2016. Each cell indicates the number of couples in which the husband and wife fall into the corresponding education categories. Rows represent the husband's education and columns represent the wife's. Education categories are abbreviated as: LHS = Less than high school; HS = High school graduate; SC = Some college; C = College degree or higher.

Table 9. Average Educational Cosine Similarity Scores Across the Full Sample, 2004-2016.

Educational attainment	LHS	HS	SC	C
LHS	0.989	0.859	0.312	0.055
HS	0.864	0.927	0.483	0.167
SC	0.193	0.447	0.917	0.846
C	0.029	0.159	0.729	0.992

Note: This table reports the average pairwise cosine similarity scores between education categories, calculated over all survey years from 2004 to 2016. Similarity values reflect the alignment of educational attributes between men (rows) and women (columns), based on gender-specific vectors derived from MEPS data. Educational categories are abbreviated as follows: LHS = Less than high school; HS = High school graduate; SC = Some college; C = College degree or higher.

Appendix

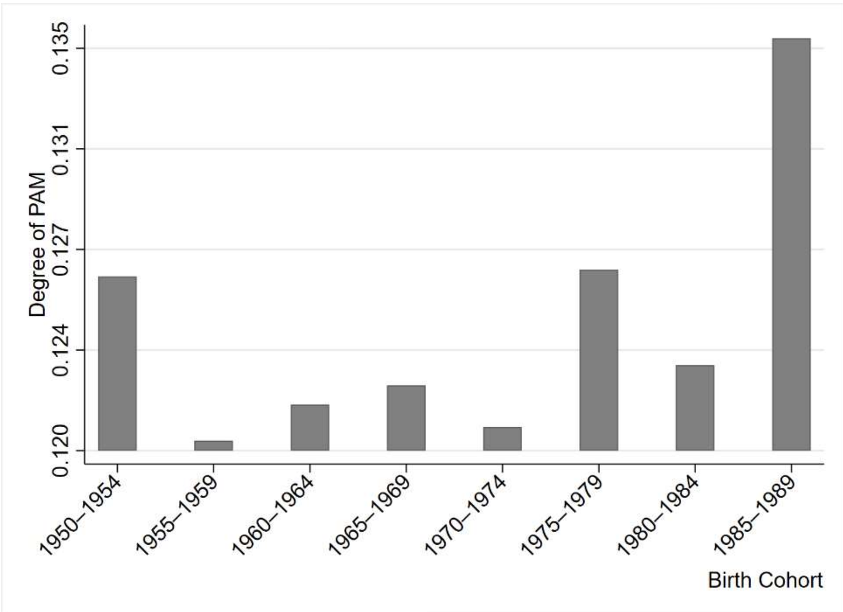
Appendix A. Occupational Assortative Mating

This appendix presents supplementary materials related to the analysis of occupational PAM discussed in the main text.

A.1. Results Based on the Identity Similarity Matrix

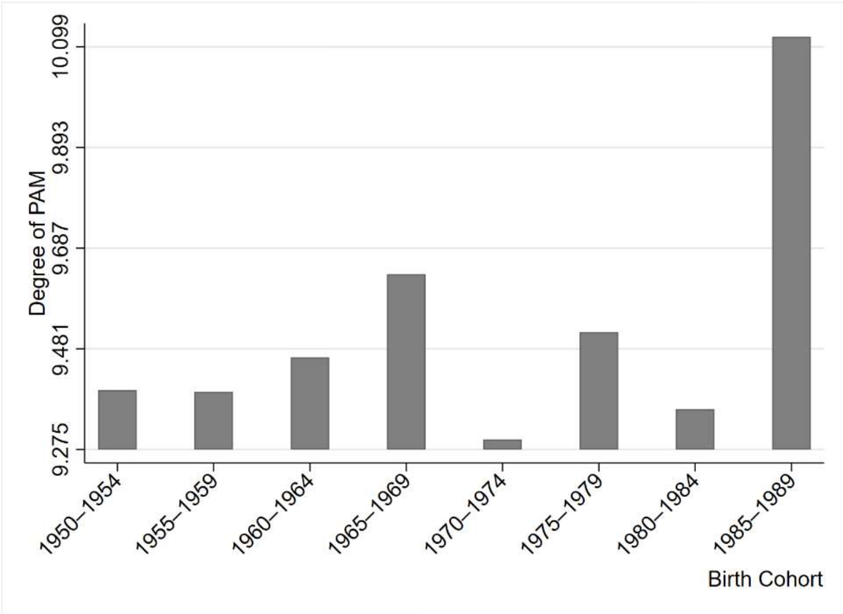
This section reports results for occupational PAM under the conventional framework, where the similarity matrix is assumed to be the identity matrix. Under this approach, perfect similarity is assigned only to identical occupational pairings, while all non-identical pairings are treated as completely dissimilar.

Figure A.1.1. Trends in Occupational PAM across Birth Cohorts Using the Perfect-Random Normalization Measure under the Identity Similarity Matrix



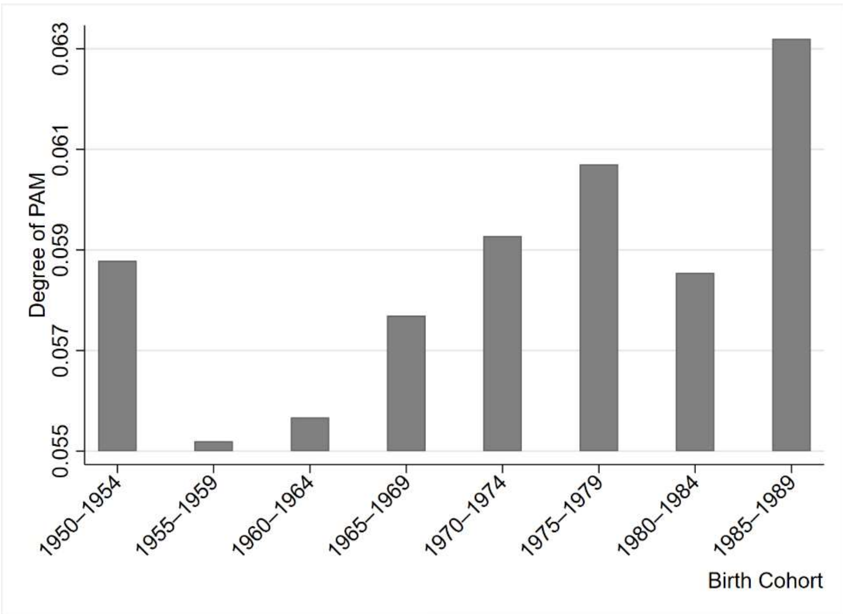
Note: This figure presents occupational PAM across five-year birth cohorts, measured using the perfect-random normalization approach under the identity similarity matrix. Similarity is defined only for exact occupational matches. For comparison, see Figure 2, which uses cosine similarity based on occupational attributes.

Figure A.1.2. Trends in Occupational PAM across Birth Cohorts Using the Aggregate Likelihood Ratio Measure under the Identity Similarity Matrix



Note: This figure presents occupational PAM across five-year birth cohorts, measured using the aggregate likelihood ratio approach under the identity similarity matrix. Similarity is defined only for exact occupational matches. For comparison, see Figure 3, which uses cosine similarity based on occupational attributes.

Figure A.1.3. Trends in Occupational PAM across Birth Cohorts Using the Weighted Similarity and the Normalized Trace Measures under the Identity Similarity Matrix



Note: This figure presents occupational PAM across five-year birth cohorts, measured using either the normalized trace or weighted similarity approach under the identity similarity matrix, which are equivalent when similarity is defined only for exact occupational matches. For comparison, see Figures 4 and 5, which uses cosine similarity based on occupational attributes.

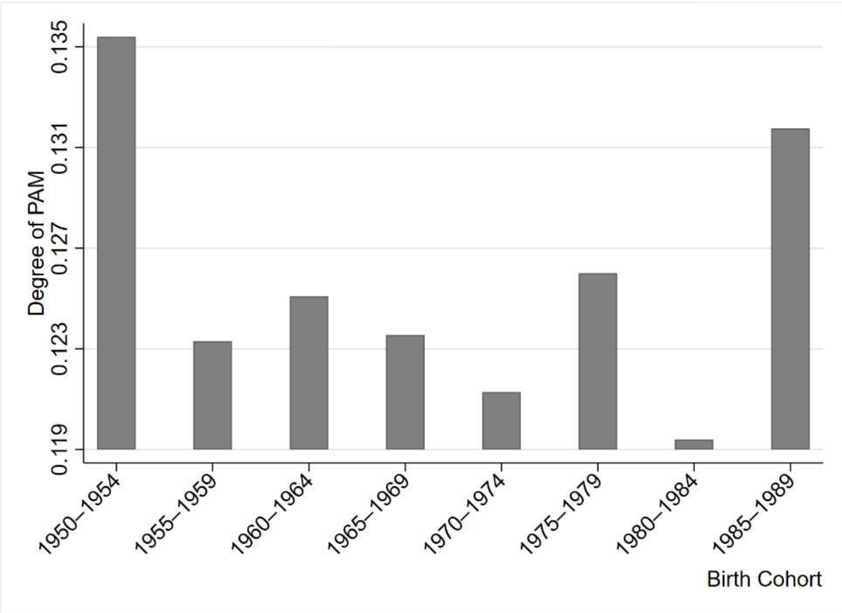
A.2. Results Using the Euclidean Similarity Matrix

Figures A.2.1 through A.2.4 present cohort patterns of PAM based on Euclidean similarity, using the four similarity-based measures described in the main text. The annual average self-similarity values (s_{ii}) for the Euclidean measure range from 0.810 to 0.865, with an overall mean of 0.840 across all years. The average pairwise occupational similarity is 0.156, with a standard deviation of 0.0374. The minimum similarity score is 0.073, and the maximum is 1. Over the sample period, the annual average similarity fluctuates narrowly between 0.155 and 0.159, and the standard deviation shows little variation —suggesting that overall structure of occupational similarity has remained stable over time. The distribution of occupational similarity scores is presented in Table A.2.1.

Figure A.2.1, which uses the perfect-random normalization measure, reveals a cohort pattern distinct from that based on cosine similarity. In particular, the PAM estimate for the 1950–1954 cohort appears substantially higher than that for any other cohort. Two factors may account for this divergence. First, earlier cohorts include a larger share of homemakers, who are excluded from our analysis due to the lack of occupational attribute data for non-market roles in the O*NET database. Because homemakers likely possess skill profiles that differ substantially from those of labor force participants, their exclusion may lead to an upward bias in the estimated PAM level. Second, couples from earlier cohorts are observed at older ages, by which time their occupational skill profiles may have become more pronounced. This deepening of specialization could inflate the estimated Euclidean distance between spouses' occupations, thereby affecting the similarity-based PAM measure.

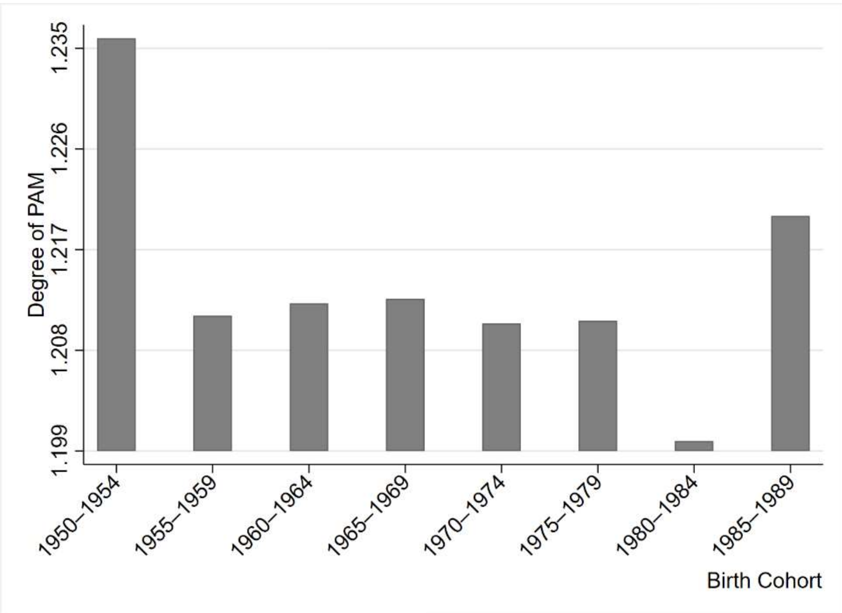
The cohort patterns observed in the remaining three figures—based on the aggregate likelihood ratio, weighted similarity, and normalized trace—are broadly similar to those in Figure A.2.1, reinforcing the distinctiveness of the Euclidean-based estimates relative to cosine-based measures.

Figure A.2.1. Trends in Occupational PAM by Birth Cohort Using the Perfect-Random Normalization Measure under the Euclidean Similarity Matrix



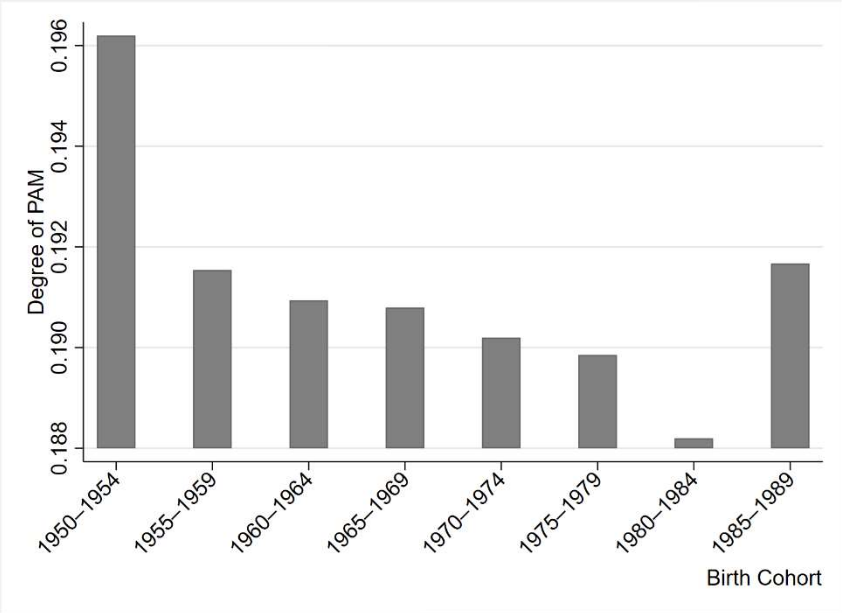
Note: This figure presents occupational PAM by five-year birth cohort using the perfect-random normalization measure, based on Euclidean similarity. For comparison, see the cosine similarity-based results in Figure 2.

Figure A.2.2. Trends in Occupational PAM by Birth Cohort Using the Aggregate Likelihood Ratio Measure under the Euclidean Similarity Matrix



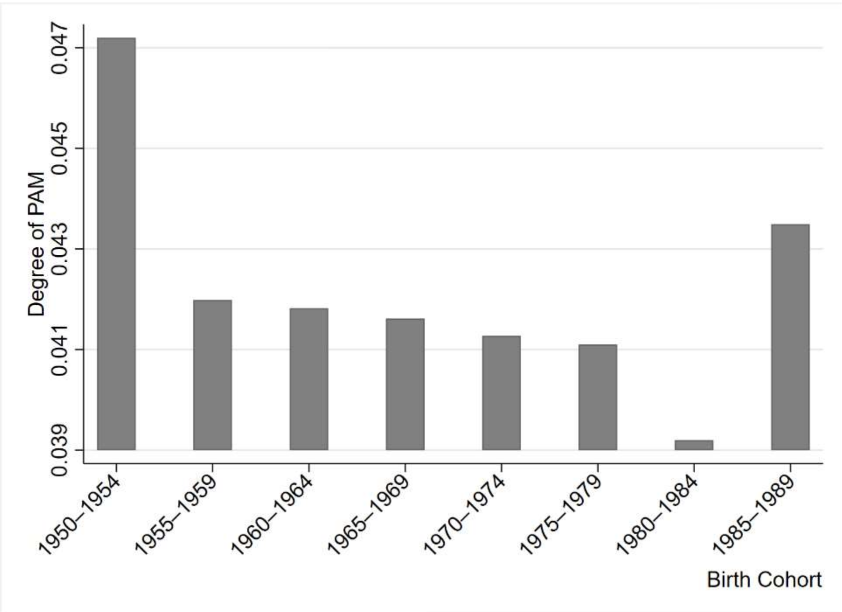
Note: This figure presents occupational PAM by five-year birth cohort using the aggregate likelihood ratio measure, based on Euclidean similarity. For comparison, see the cosine similarity-based results in Figure 3.

Figure A.2.3. Trends in Occupational PAM by Birth Cohort Using the Weighted Similarity Measure under the Euclidean Similarity Matrix



Note: This figure presents occupational PAM by five-year birth cohort using the weighted similarity measure, based on Euclidean similarity. For comparison, see the cosine similarity-based results in Figure 4.

Figure A.2.4. Trends in Occupational PAM by Birth Cohort Using the Normalized Trace Measure under the Euclidean Similarity Matrix



Note: This figure presents occupational PAM by five-year birth cohort using the normalized trace measure, based on Euclidean similarity. For comparison, see the cosine similarity-based results in Figure 5.

Table A.2.1. Distribution of Pairwise Occupational Euclidean Similarity Scores

Range	Share (%)
s = 1	0.09
0.9 ≤ s < 1	0.01
0.8 ≤ s < 0.9	0.01
0.7 ≤ s < 0.8	0.01
0.6 ≤ s < 0.7	0.02
0.5 ≤ s < 0.6	0.01
0.4 ≤ s < 0.5	0.01
0.3 ≤ s < 0.4	0.03
0.2 ≤ s < 0.3	4.60
0.15 ≤ s < 0.2	48.44
0.10 ≤ s < 0.15	46.49
s < 0.10	0.28

Mean Similarity: 0.156 **Standard Deviation:** 0.0374

A.3. List of Occupational Attributes

This section documents the occupational attributes used to construct the similarity matrices underlying our measures of occupational PAM. The attributes are drawn from the O*NET database and span key domains including abilities, skills, knowledge, work styles, and work values. For each attribute, we report its variable name and a brief description following the O*NET classification system.

The final set includes 148 attributes, which form the basis for calculating both cosine and Euclidean similarity between occupations. These standardized variables capture a wide range of job-relevant characteristics and are essential to quantifying occupational proximity in our analysis.

Table A.3.1. Occupational Attributes in O*NET Data

Area	**Element Name**	**Description**
Abilities (ABs) are enduring attributes of the individual that influence performance.		
AB	Oral Comprehension	The ability to listen to and understand information and ideas presented through spoken words and sentences.
AB	Written Comprehension	The ability to read and understand information and ideas presented in writing.
AB	Oral Expression	The ability to communicate information and ideas in speaking so others will understand.

AB	Written Expression	The ability to communicate information and ideas in writing so others will understand.
AB	Fluency of Ideas	The ability to come up with a number of ideas about a topic (the number of ideas is important, not their quality, correctness, or creativity).
AB	Originality	The ability to come up with unusual or clever ideas about a given topic or situation, or to develop creative ways to solve a problem.
AB	Problem Sensitivity	The ability to tell when something is wrong or is likely to go wrong. It does not involve solving the problem, only recognizing that there is a problem.
AB	Deductive Reasoning	The ability to apply general rules to specific problems to produce answers that make sense.
AB	Inductive Reasoning	The ability to combine pieces of information to form general rules or conclusions (includes finding a relationship among seemingly unrelated events).
AB	Information Ordering	The ability to arrange things or actions in a certain order or pattern according to a specific rule or set of rules (e.g., patterns of numbers, letters, words, pictures, mathematical operations).
AB	Category Flexibility	The ability to generate or use different sets of rules for combining or grouping things in different ways.
AB	Mathematical Reasoning	The ability to choose the right mathematical methods or formulas to solve a problem.
AB	Number Facility	The ability to add, subtract, multiply, or divide quickly and correctly.
AB	Memorization	The ability to remember information such as words, numbers, pictures, and procedures.
AB	Speed of Closure	The ability to quickly make sense of, combine, and organize information into meaningful patterns.
AB	Flexibility of Closure	The ability to identify or detect a known pattern (a figure, object, word, or sound) that is hidden in other distracting material.
AB	Perceptual Speed	The ability to quickly and accurately compare similarities and differences among sets of letters, numbers, objects, pictures, or patterns. The things to be compared may be presented at the same time or one after the other. This ability also includes comparing a presented object with a remembered object.
AB	Spatial Orientation	The ability to know your location in relation to the environment or to know where other objects are in relation to you.
AB	Visualization	The ability to imagine how something will look after it is moved around or when its parts are moved or rearranged.
AB	Selective Attention	The ability to concentrate on a task over a period of time without being distracted.
AB	Time Sharing	The ability to shift back and forth between two or more activities or sources of information (such as speech, sounds, touch, or other sources).
AB	Arm-Hand Steadiness	The ability to keep your hand and arm steady while moving your arm or while holding your arm and hand in one position.
AB	Manual Dexterity	The ability to quickly move your hand, your hand together with your arm, or your two hands to grasp, manipulate, or assemble objects.

AB	Finger Dexterity	The ability to make precisely coordinated movements of the fingers of one or both hands to grasp, manipulate, or assemble very small objects.
AB	Control Precision	The ability to quickly and repeatedly adjust the controls of a machine or a vehicle to exact positions.
AB	Multi-limb Coordination	The ability to coordinate two or more limbs (for example, two arms, two legs, or one leg and one arm) while sitting, standing, or lying down. It does not involve performing the activities while the whole body is in motion.
AB	Response Orientation	The ability to choose quickly between two or more movements in response to two or more different signals (lights, sounds, pictures). It includes the speed with which the correct response is started with the hand, foot, or other body part.
AB	Rate Control	The ability to time your movements or the movement of a piece of equipment in anticipation of changes in the speed and/or direction of a moving object or scene.
AB	Reaction Time	The ability to quickly respond (with the hand, finger, or foot) to a signal (sound, light, picture) when it appears.
AB	Wrist-Finger Speed	The ability to make fast, simple, repeated movements of the fingers, hands, and wrists.
AB	Speed of Limb Movement	The ability to quickly move the arms and legs.
AB	Static Strength	The ability to exert maximum muscle force to lift, push, pull, or carry objects.
AB	Explosive Strength	The ability to use short bursts of muscle force to propel oneself (as in jumping or sprinting), or to throw an object.
AB	Dynamic Strength	The ability to exert muscle force repeatedly or continuously over time. This involves muscular endurance and resistance to muscle fatigue.
AB	Trunk Strength	The ability to use your abdominal and lower back muscles to support part of the body repeatedly or continuously over time without "giving out" or fatiguing.
AB	Stamina	The ability to exert yourself physically over long periods of time without getting winded or out of breath.
AB	Extent Flexibility	The ability to bend, stretch, twist, or reach with your body, arms, and/or legs.
AB	Dynamic Flexibility	The ability to quickly and repeatedly bend, stretch, twist, or reach out with your body, arms, and/or legs.
AB	Gross Body Coordination	The ability to coordinate the movement of your arms, legs, and torso together when the whole body is in motion.
AB	Gross Body Equilibrium	The ability to keep or regain your body balance or stay upright when in an unstable position.
AB	Near Vision	The ability to see details at close range (within a few feet of the observer).
AB	Far Vision	The ability to see details at a distance.
AB	Visual Color Discrimination	The ability to match or detect differences between colors, including shades of color and brightness.
AB	Night Vision	The ability to see under low-light conditions.
AB	Peripheral Vision	The ability to see objects or movement of objects to one's side when the eyes are looking ahead.
AB	Depth Perception	The ability to judge which of several objects is closer or farther away from you, or to judge the distance between you and an object.

AB	Glare Sensitivity	The ability to see objects in the presence of a glare or bright lighting.
AB	Hearing Sensitivity	The ability to detect or tell the differences between sounds that vary in pitch and loudness.
AB	Auditory Attention	The ability to focus on a single source of sound in the presence of other distracting sounds.
AB	Sound Localization	The ability to tell the direction from which a sound originated.
AB	Speech Recognition	The ability to identify and understand the speech of another person.
AB	Speech Clarity	The ability to speak clearly so others can understand you.
Occupational Interests (OIs) are preferences for work environments and outcomes.		
OI	Realistic	Work involves designing, building, or repairing of equipment, materials, or structures, engaging in physical activity, or working outdoors. Realistic occupations are often associated with engineering, mechanics and electronics, construction, woodworking, transportation, machine operation, agriculture, animal services, physical or manual labor, athletics, or protective services.
OI	Investigative	Work involves studying and researching non-living objects, living organisms, disease or other forms of impairment, or human behavior. Investigative occupations are often associated with physical, life, medical, or social sciences, and can be found in the fields of humanities, mathematics/statistics, information technology, or health care service.
OI	Artistic	Work involves creating original visual artwork, performances, written works, food, or music for a variety of media, or applying artistic principles to the design of various objects and materials. Artistic occupations are often associated with visual arts, applied arts and design, performing arts, music, creative writing, media, or culinary art.
OI	Social	Work involves helping, teaching, advising, assisting, or providing service to others. Social occupations are often associated with social, health care, personal service, teaching/education, or religious activities.
OI	Enterprising	Work involves managing, negotiating, marketing, or selling, typically in a business setting, or leading or advising people in political and legal situations. Enterprising occupations are often associated with business initiatives, sales, marketing/advertising, finance, management/administration, professional advising, public speaking, politics, or law.
OI	Conventional	Work involves following procedures and regulations to organize information or data, typically in a business setting. Conventional occupations are often associated with office work, accounting, mathematics/statistics, information technology, finance, or human resources.
Knowledge (KN) is organized sets of principles and facts applying in general domains.		
KN	Administration and Management	Knowledge of business and management principles involved in strategic planning, resource allocation, human resources modeling, leadership technique, production methods, and coordination of people and resources.

KN	Administrative	Knowledge of administrative and office procedures and systems such as word processing, managing files and records, stenography and transcription, designing forms, and workplace terminology.
KN	Economics and Accounting	Knowledge of economic and accounting principles and practices, the financial markets, banking, and the analysis and reporting of financial data.
KN	Sales and Marketing	Knowledge of principles and methods for showing, promoting, and selling products or services. This includes marketing strategy and tactics, product demonstration, sales techniques, and sales control systems.
KN	Customer and Personal Service	Knowledge of principles and processes for providing customer and personal services. This includes customer needs assessment, meeting quality standards for services, and evaluation of customer satisfaction.
KN	Personnel and Human Resources	Knowledge of principles and procedures for personnel recruitment, selection, training, compensation and benefits, labor relations and negotiation, and personnel information systems.
KN	Production and Processing	Knowledge of raw materials, production processes, quality control, costs, and other techniques for maximizing the effective manufacture and distribution of goods.
KN	Food Production	Knowledge of techniques and equipment for planting, growing, and harvesting food products (both plant and animal) for consumption, including storage/handling techniques.
KN	Computers and Electronics	Knowledge of circuit boards, processors, chips, electronic equipment, and computer hardware and software, including applications and programming.
KN	Engineering and Technology	Knowledge of the practical application of engineering science and technology. This includes applying principles, techniques, procedures, and equipment to the design and production of various goods and services.
KN	Design	Knowledge of design techniques, tools, and principles involved in production of precision technical plans, blueprints, drawings, and models.
KN	Building and Construction	Knowledge of materials, methods, and the tools involved in the construction or repair of houses, buildings, or other structures such as highways and roads.
KN	Mechanical	Knowledge of machines and tools, including their designs, uses, repair, and maintenance.
KN	Mathematics	Knowledge of arithmetic, algebra, geometry, calculus, statistics, and their applications.
KN	Physics	Knowledge and prediction of physical principles, laws, their interrelationships, and applications to understanding fluid, material, and atmospheric dynamics, and mechanical, electrical, atomic and sub-atomic structures and processes.
KN	Chemistry	Knowledge of the chemical composition, structure, and properties of substances and of the chemical processes and transformations that they undergo. This includes uses of chemicals and their interactions, danger signs,

		production techniques, and disposal methods.
KN	Biology	Knowledge of plant and animal organisms, their tissues, cells, functions, interdependencies, and interactions with each other and the environment.
KN	Psychology	Knowledge of human behavior and performance; individual differences in ability, personality, and interests; learning and motivation; psychological research methods; and the assessment and treatment of behavioral and affective disorders.
KN	Sociology and Anthropology	Knowledge of group behavior and dynamics, societal trends and influences, human migrations, ethnicity, cultures, and their history and origins.
KN	Geography	Knowledge of principles and methods for describing the features of land, sea, and air masses, including their physical characteristics, locations, interrelationships, and distribution of plant, animal, and human life.
KN	Medicine and Dentistry	Knowledge of the information and techniques needed to diagnose and treat human injuries, diseases, and deformities. This includes symptoms, treatment alternatives, drug properties and interactions, and preventive health-care measures.
KN	Therapy and Counseling	Knowledge of principles, methods, and procedures for diagnosis, treatment, and rehabilitation of physical and mental dysfunctions, and for career counseling and guidance.
KN	Education and Training	Knowledge of principles and methods for curriculum and training design, teaching and instruction for individuals and groups, and the measurement of training effects.
KN	English Language	Knowledge of the structure and content of the English language including the meaning and spelling of words, rules of composition, and grammar.
KN	Foreign Language	Knowledge of the structure and content of a foreign (non-English) language including the meaning and spelling of words, rules of composition and grammar, and pronunciation.
KN	Fine Arts	Knowledge of the theory and techniques required to compose, produce, and perform works of music, dance, visual arts, drama, and sculpture.
KN	History and Archeology	Knowledge of historical events and their causes, indicators, and effects on civilizations and cultures.
KN	Philosophy and Theology	Knowledge of different philosophical systems and religions. This includes their basic principles, values, ethics, ways of thinking, customs, practices, and their impact on human culture.
KN	Public Safety and Security	Knowledge of relevant equipment, policies, procedures, and strategies to promote effective local, state, or national security operations for the protection of people, data, property, and institutions.
KN	Law and Government	Knowledge of laws, legal codes, court procedures, precedents, government regulations, executive orders, agency rules, and the democratic political process.
KN	Telecommunications	Knowledge of transmission, broadcasting, switching, control, and operation of telecommunications systems.

KN	Communications and Media	Knowledge of media production, communication, and dissemination techniques and methods. This includes alternative ways to inform and entertain via written, oral, and visual media.
KN	Transportation	Knowledge of principles and methods for moving people or goods by air, rail, sea, or road, including the relative costs and benefits.
Skills (SKs) are developed capacities that facilitate learning or the more rapid acquisition of knowledge and facilitate performance of activities that occur across jobs.		
SK	Reading Comprehension	Understanding written sentences and paragraphs in work-related documents.
SK	Active Listening	Giving full attention to what other people are saying, taking time to understand the points being made, asking questions as appropriate, and not interrupting at inappropriate times.
SK	Writing	Communicating effectively in writing as appropriate for the needs of the audience.
SK	Speaking	Talking to others to convey information effectively.
SK	Mathematics	Using mathematics to solve problems.
SK	Science	Using scientific rules and methods to solve problems.
SK	Critical Thinking	Using logic and reasoning to identify the strengths and weaknesses of alternative solutions, conclusions, or approaches to problems.
SK	Active Learning	Understanding the implications of new information for both current and future problem-solving and decision-making.
SK	Learning Strategies	Selecting and using training/instructional methods and procedures appropriate for the situation when learning or teaching new things.
SK	Monitoring	Monitoring/Assessing performance of yourself, other individuals, or organizations to make improvements or take corrective action.
SK	Social Perceptiveness	Being aware of others' reactions and understanding why they react as they do.
SK	Coordination	Adjusting actions in relation to others' actions.
SK	Persuasion	Persuading others to change their minds or behavior.
SK	Negotiation	Bringing others together and trying to reconcile differences.
SK	Instructing	Teaching others how to do something.
SK	Service Orientation	Actively looking for ways to help people.
SK	Complex Problem Solving	Identifying complex problems and reviewing related information to develop and evaluate options and implement solutions.
SK	Operations Analysis	Analyzing needs and product requirements to create a design.
SK	Technology Design	Generating or adapting equipment and technology to serve user needs.
SK	Equipment Selection	Determining the kind of tools and equipment needed to do a job.
SK	Installation	Installing equipment, machines, wiring, or programs to meet specifications.
SK	Programming	Writing computer programs for various purposes.
SK	Operations Monitoring	Watching gauges, dials, or other indicators to make sure a machine is working properly.
SK	Operation and Control	Controlling operations of equipment or systems.

SK	Equipment Maintenance	Performing routine maintenance on equipment and determining when and what kind of maintenance is needed.
SK	Troubleshooting	Determining causes of operating errors and deciding what to do about it.
SK	Repairing	Repairing machines or systems using the needed tools.
SK	Quality Control Analysis	Conducting tests and inspections of products, services, or processes to evaluate quality or performance.
SK	Judgment and Decision Making	Considering the relative costs and benefits of potential actions to choose the most appropriate one.
SK	Systems Analysis	Determining how a system should work and how changes in conditions, operations, and the environment will affect outcomes.
SK	Systems Evaluation	Identifying measures or indicators of system performance and the actions needed to improve or correct performance, relative to the goals of the system.
SK	Time Management	Managing one's own time and the time of others.
SK	Management of Financial Resources	Determining how money will be spent to get the work done, and accounting for these expenditures.
SK	Management of Material Resources	Obtaining and seeing to the appropriate use of equipment, facilities, and materials needed to do certain work.
SK	Management of Personnel Resources	Motivating, developing, and directing people as they work, identifying the best people for the job.
Work Styles (WSs) are personal characteristics that can affect how well someone performs a job.		
WS	Achievement/Effort	Job requires establishing and maintaining personally challenging achievement goals and exerting effort toward mastering tasks.
WS	Persistence	Job requires persistence in the face of obstacles.
WS	Initiative	Job requires a willingness to take on responsibilities and challenges.
WS	Leadership	Job requires a willingness to lead, take charge, and offer opinions and direction.
WS	Cooperation	Job requires being pleasant with others on the job and displaying a good-natured, cooperative attitude.
WS	Concern for Others	Job requires being sensitive to others' needs and feelings and being understanding and helpful on the job.
WS	Social Orientation	Job requires preferring to work with others rather than alone, and being personally connected with others on the job.
WS	Self-Control	Job requires maintaining composure, keeping emotions in check, controlling anger, and avoiding aggressive behavior, even in very difficult situations.
WS	Stress Tolerance	Job requires accepting criticism and dealing calmly and effectively with high-stress situations.
WS	Adaptability/Flexibility	Job requires being open to change (positive or negative) and to considerable variety in the workplace.
WS	Dependability	Job requires being reliable, responsible, and dependable, and fulfilling obligations.
WS	Attention to Detail	Job requires being careful about detail and thorough in completing work tasks.
WS	Integrity	Job requires being honest and ethical.

WS	Independence	Job requires developing one's own ways of doing things, guiding oneself with little or no supervision, and depending on oneself to get things done.
WS	Innovation	Job requires creativity and alternative thinking to develop new ideas for and answers to work-related problems.
WS	Analytical Thinking	Job requires analyzing information and using logic to address work-related issues and problems.
Work Values (WVs) are global aspects of work that are important to a person's satisfaction.		
WV	Achievement	Occupations that satisfy this work value are results oriented and allow employees to use their strongest abilities, giving them a feeling of accomplishment.
WV	Independence	Occupations that satisfy this work value allow employees to work on their own and make decisions.
WV	Recognition	Occupations that satisfy this work value offer advancement, potential for leadership, and are often considered prestigious.
WV	Relationships	Occupations that satisfy this work value allow employees to provide service to others and work with co-workers in a friendly non-competitive environment.
WV	Support	Occupations that satisfy this work value offer supportive management that stands behind employees.
WV	Working Conditions	Occupations that satisfy this work value offer job security and good working conditions. Corresponding needs are activities, compensation, security, and variety.

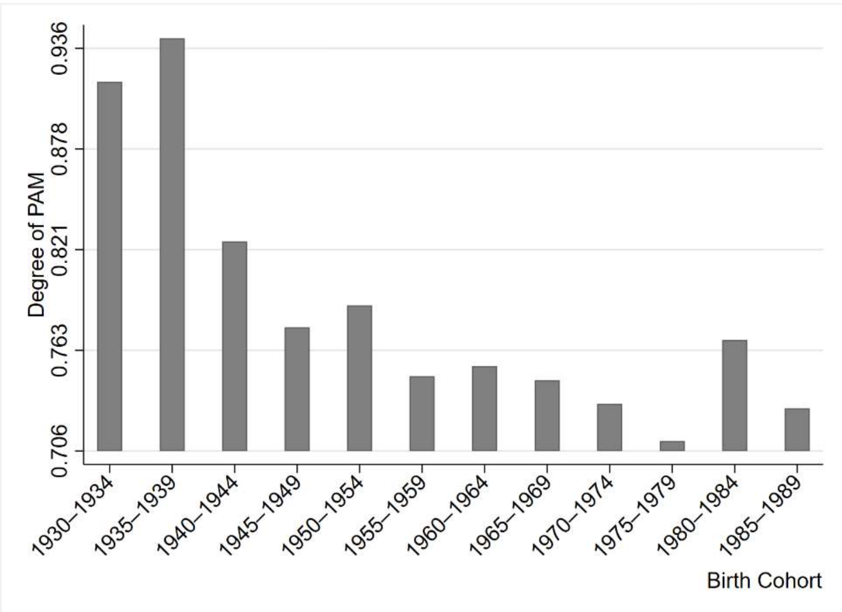
Appendix B. Religious PAM

This appendix provides supplementary materials for the analysis of religious PAM discussed in the main text.

B.1. Results Using the Identity Similarity Matrix

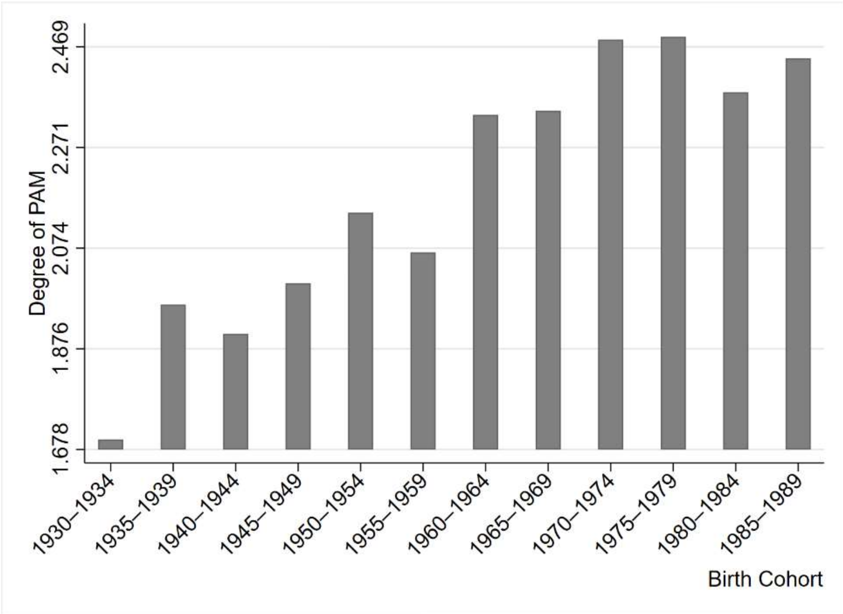
This section presents benchmark estimates of religious PAM under the conventional assumption that the similarity matrix is defined as an identity matrix—that is, only exact matches between religions affiliations are considered similar, while all other pairings are treated as entirely dissimilar.

Figure B.1.1. Trends in Religious PAM by Birth Cohort Using the Perfect-Random Normalization Measure under the Identity Similarity Matrix



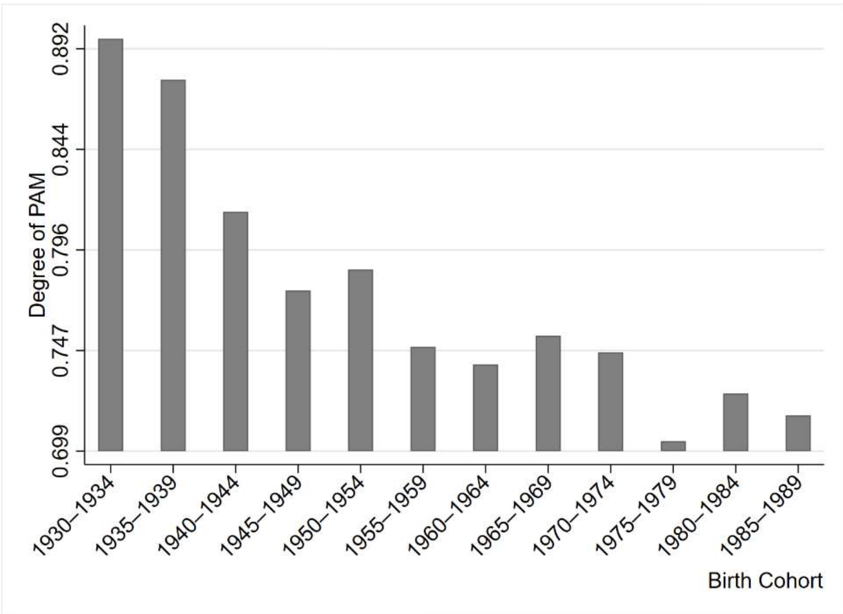
Note: This figure presents religious PAM across five-year birth cohorts, measured using the perfect-random normalization approach under the identity similarity matrix. Similarity is defined only for exact religious matches. For comparison, see Figure 7, which uses cosine similarity.

Figure B.1.2. Trends in Religious PAM by Birth Cohort Using the Aggregate Likelihood Ratio Measure under the Identity Similarity Matrix



Note: This figure presents religious PAM across five-year birth cohorts, measured using the aggregate likelihood ratio approach under the identity similarity matrix. Similarity is defined only for exact religious matches. For comparison, see Figure 8, which uses cosine similarity.

Figure B.1.3. Trends in Religious PAM by Birth Cohort Using the Weighted Similarity and the Normalized Trace Measures under the Identity Similarity Matrix



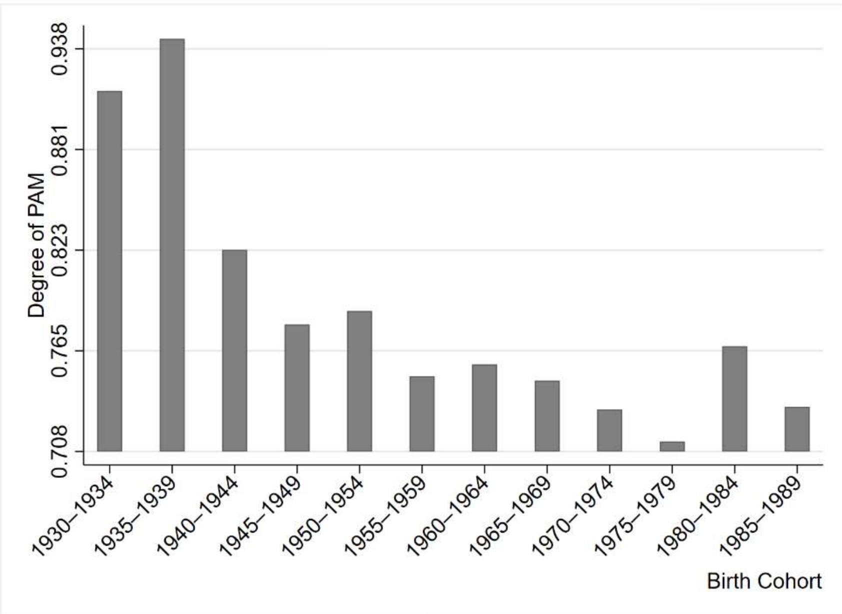
Note: Religious PAM is measured across cohorts using either the normalized trace or the weighted similarity measure, both of which yield identical results under the assumption of perfect similarity only between identical religious denominations. This figure serves as a benchmark for comparison with the cosine similarity-based results presented in Figures 9 and 10.

B.2. Results Using the Euclidean Similarity Matrix

Figures B.2.1 through B.2.4 present cohort patterns of religious PAM based on Euclidean similarity, using the four similarity-based measures described in the main text. By construction, each category's similarity with itself is normalized to 1, so the overall average self-similarity (sii) for the Euclidean measure is exactly 1 in every year. The average pairwise religious similarity is 0.264, with a standard deviation of 0.262. The religious Euclidean similarity matrix averaged over the entire sample is presented in Table B.2.1.

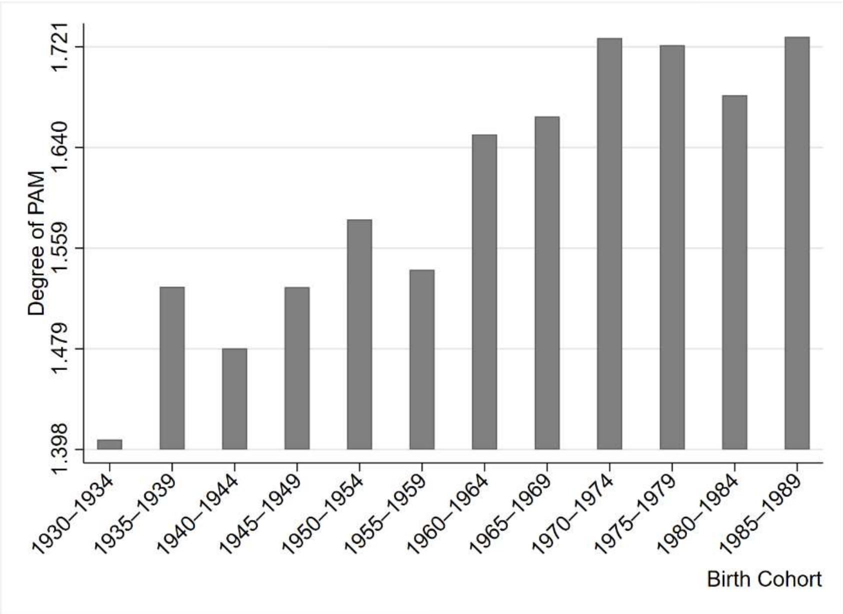
The cohort patterns observed in Figures B.2.1 through B.2.4 are broadly similar to those in Figures 7 through 10 in the main text which are based on the cosine similarity matrix.

Figure B.2.1. Trends in Religious PAM by Birth Cohort Using the Perfect-Random Normalization Measure under the Euclidean Similarity Matrix



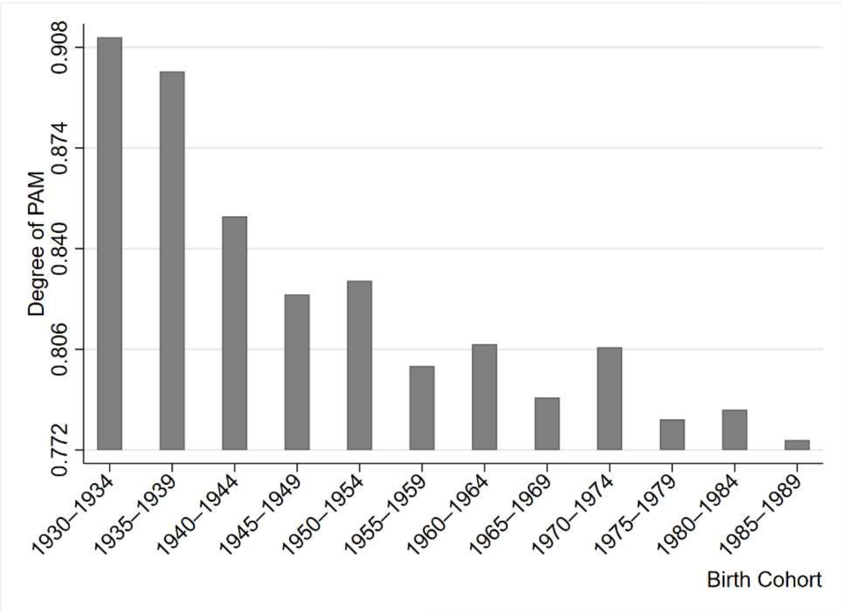
Note: This figure presents religious PAM by five-year birth cohort using the perfect-random normalization measure, based on Euclidean similarity. For comparison, see the cosine similarity-based results in Figure 7.

Figure B.2.2. Trends in Religious PAM by Birth Cohort Using the Aggregate Likelihood Ratio Measure under the Euclidean Similarity Matrix



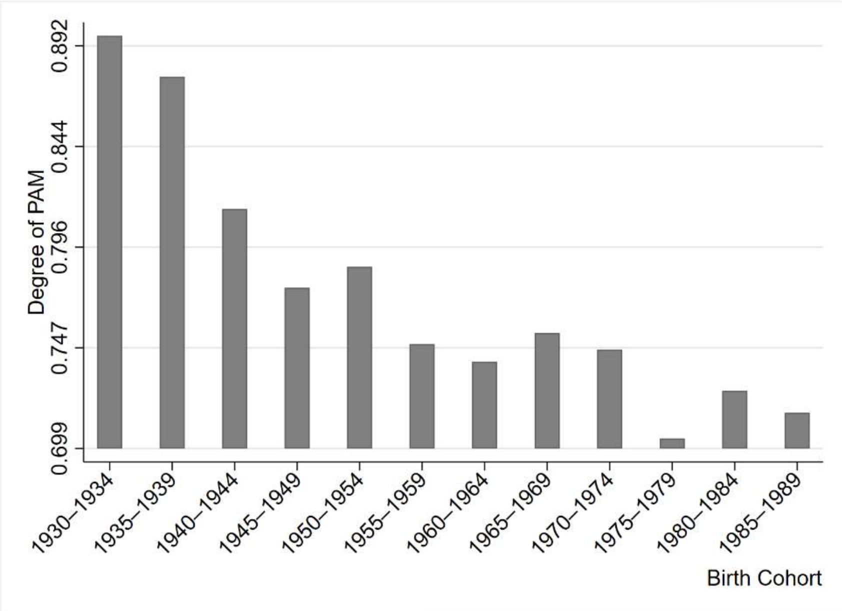
Note: This figure presents religious PAM by five-year birth cohort using the aggregate likelihood ratio measure, based on Euclidean similarity. For comparison, see the cosine similarity-based results in Figure 8.

Figure B.2.3. Trends in Religious PAM by Birth Cohort Using the Weighted Similarity Measure under the Euclidean Similarity Matrix



Note: This figure presents religious PAM by five-year birth cohort using the weighted similarity measure, based on Euclidean similarity. For comparison, see the cosine similarity-based results in Figure 9.

Figure B.2.4. Trends in Religious PAM by Birth Cohort Using the Normalized Trace Measure under the Euclidean Similarity Matrix



Note: This figure presents religious PAM by five-year birth cohort using the normalized trace measure, based on Euclidean similarity. For comparison, see the cosine similarity-based results in Figure 10.

Table B.2.1. Average Religious Euclidean Similarity Scores Across the Full Sample, 2004–2022

	None	Catholic	Protestant	Orthodox	Jew	Muslim	Hindu	Buddhist	Other Christian	Other
None	1.000	0.234	0.191	0.126	0.182	0.122	0.123	0.151	0.186	0.211
Catholic	0.234	1.000	0.343	0.128	0.187	0.124	0.125	0.153	0.190	0.211
Protestant	0.191	0.343	1.000	0.124	0.176	0.123	0.121	0.147	0.186	0.214
Orthodox	0.126	0.128	0.124	1.000	0.112	0.097	0.100	0.108	0.118	0.130
Jew	0.182	0.187	0.176	0.112	1.000	0.113	0.117	0.136	0.152	0.157
Muslim	0.122	0.124	0.123	0.097	0.113	1.000	0.094	0.104	0.115	0.118
Hindu	0.123	0.125	0.121	0.100	0.117	0.094	1.000	0.107	0.116	0.121
Buddhist	0.151	0.153	0.147	0.108	0.136	0.104	0.107	1.000	0.140	0.138
Other Christian	0.186	0.190	0.186	0.118	0.152	0.115	0.116	0.140	1.000	0.173
Other	0.211	0.211	0.214	0.130	0.157	0.118	0.121	0.138	0.173	1.000

Note: This table reports average pairwise Euclidean similarity scores between religious categories, calculated across the full GSS sample from 2004 to 2022. Each cell represents the mean similarity between the corresponding religious affiliations (for comparison, see Table 7).

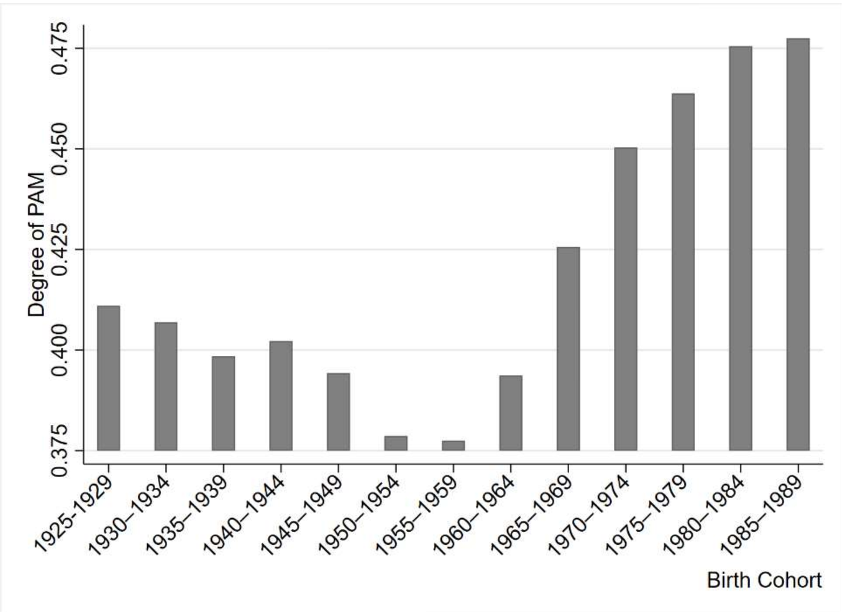
Appendix C. Educational PAM

This appendix provides supplementary materials for the analysis of educational PAM discussed in the main text.

Appendix C.1. Results Using the Identity Similarity Matrix

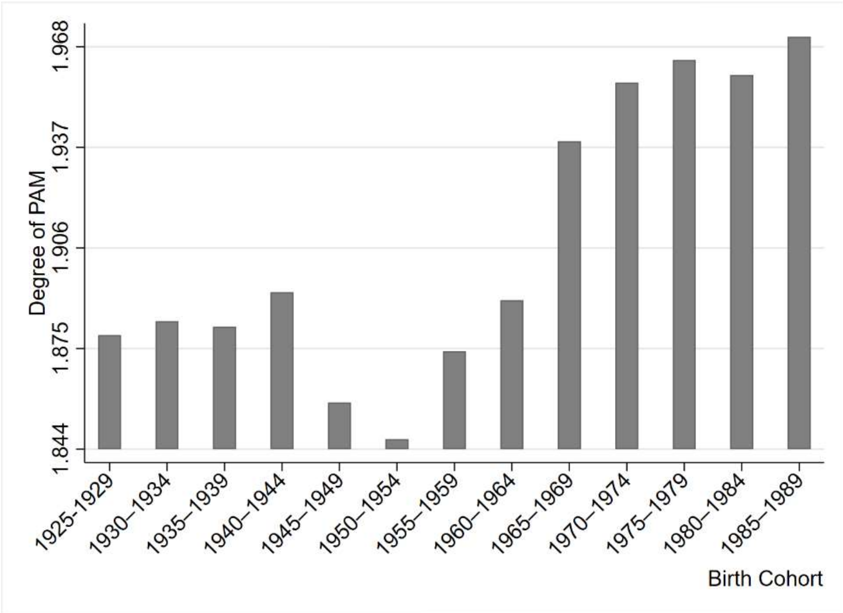
This section reports results for educational PAM under the conventional approach, where the similarity matrix is assumed to be the identity matrix. Under this approach, perfect similarity is assigned only to identical educational pairings, while all non-identical pairings are treated as completely dissimilar.

Figure C.1.1. Trends in Educational PAM across Birth Cohorts Using the Perfect-Random Normalization Measure under the Identity Similarity Matrix



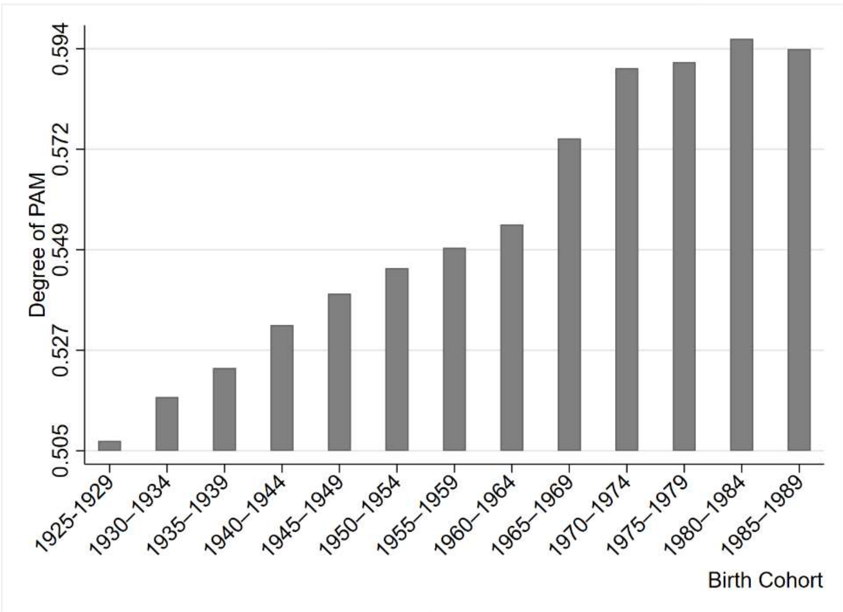
Note: This figure presents educational PAM across five-year birth cohorts, measured using the perfect-random normalization approach. Similarity is defined only for exact educational matches. For comparison, see Figure 12, which uses our similarity-based approach.

Figure C.1.2. Trends in Educational PAM across Birth Cohorts Using the Aggregate Likelihood Ratio Measure under the Identity Similarity Matrix



Note: This figure presents educational PAM across five-year birth cohorts, measured using the aggregate likelihood ratio approach. Similarity is defined only for exact educational matches. For comparison, see Figure 13, which uses our similarity-based approach.

Figure C.1.3. Trends in Educational PAM across Birth Cohorts Using the Weighted Similarity and Normalized Trace Measures under the Identity Similarity Matrix



Note: Educational PAM is measured by using either the normalized trace or the weighted similarity measure, assuming perfect similarity only between identical educational levels. For comparison, see Figure 14 and 15.

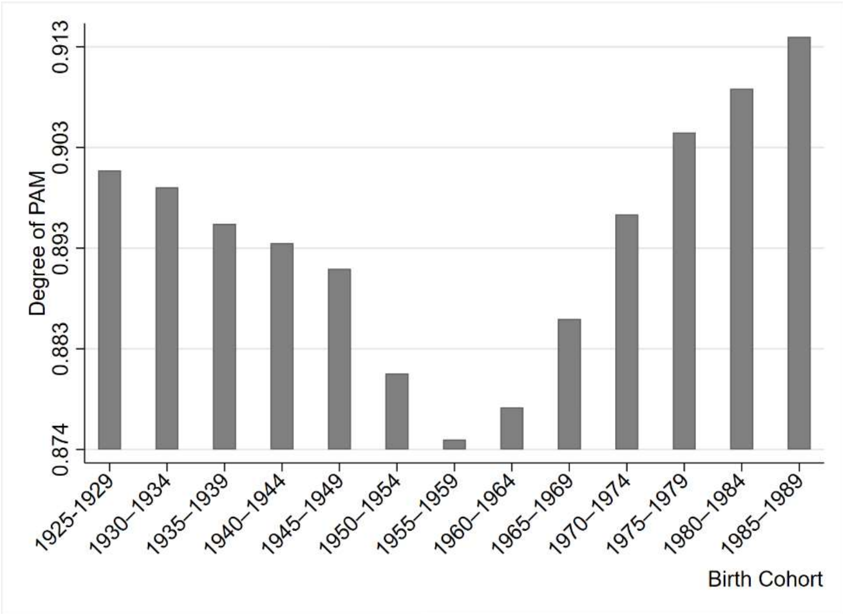
C.2. Results Using the Euclidean Similarity Matrix

Figures C.2.1 through C.2.4 present cohort-level trends in educational PAM based on Euclidean similarity, using the four similarity-based measures discussed in the main text. The average self-similarity value (s_{ii}) under Euclidean similarity is 0.150 across all years. The mean pairwise educational similarity is 0.145, with a standard deviation of 0.024. Table C.2.1 displays the full educational Euclidean similarity matrix averaged over the entire sample.

The cohort patterns shown in Figures C.2.1 through C.2.4 differ notably from those in Figures 12 through 15, which are based on cosine similarity. In particular, PAM trends under Euclidean similarity vary more across the four measures, whereas the cosine-based measures exhibit relatively consistent cohort patterns.

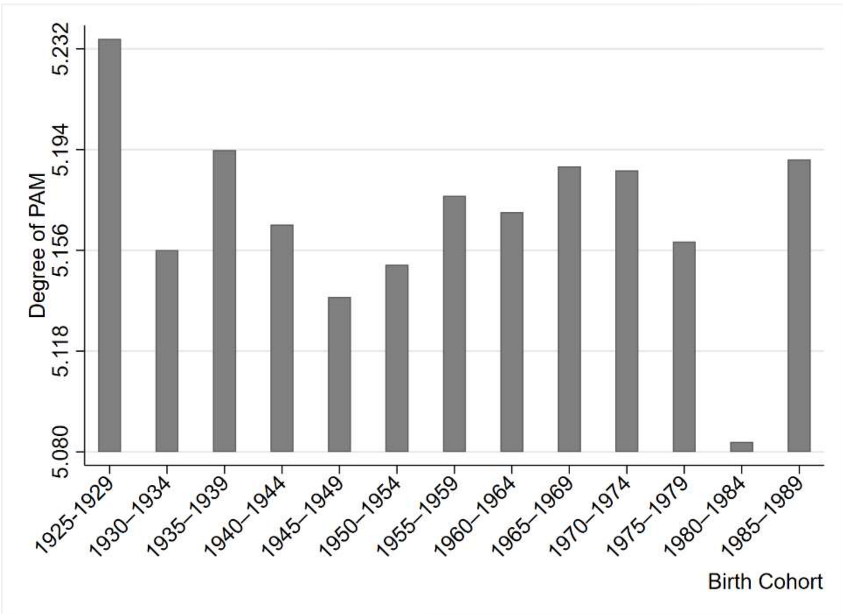
Nonetheless, a common feature emerges across Figures C.2.1 through C.2.4: earlier cohorts display relatively high PAM scores compared to the cosine-based results in Figures 12 through 15. This mirrors the pattern observed in occupational PAM using Euclidean similarity, where the 1950–1954 cohort exhibited an unusually high PAM estimate. As discussed in Appendix A.2, this result may reflect the exclusion of homemakers—who lack occupational attributes in O*NET—and the fact that older cohorts appear in the data at later stages in life, when their educational or occupational profiles may be more established. These factors may likewise contribute to elevated PAM scores among earlier cohorts in the educational domain when using Euclidean similarity.

Figure C.2.1. Trends in Educational PAM by Birth Cohort Using the Perfect-Random Normalization Measure under the Euclidean Similarity Matrix



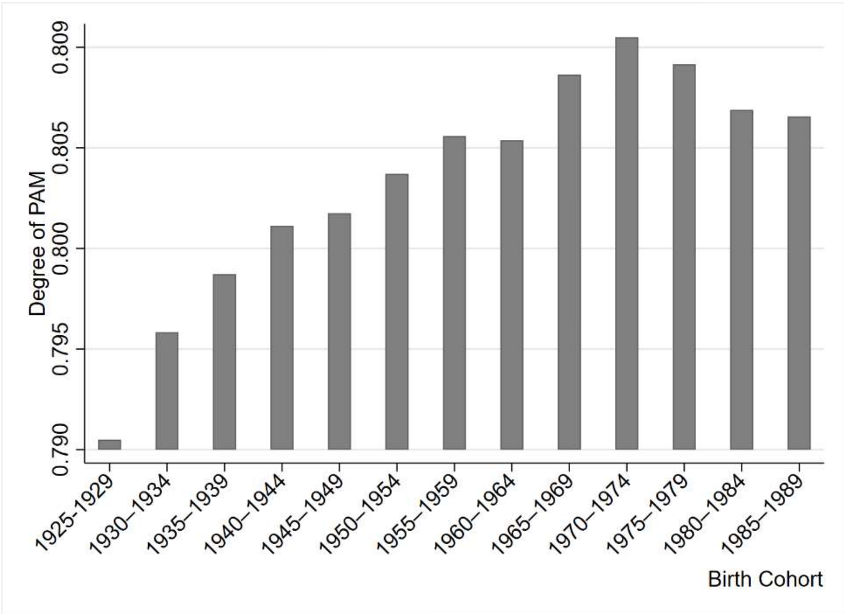
Note: This figure presents educational PAM by five-year birth cohort using the perfect-random normalization measure, based on Euclidean similarity. For comparison, see the cosine similarity-based results in Figure 12.

Figure C.2.2. Trends in Educational PAM by Birth Cohort Using the Aggregate Likelihood Ratio Measure under the Euclidean Similarity Matrix



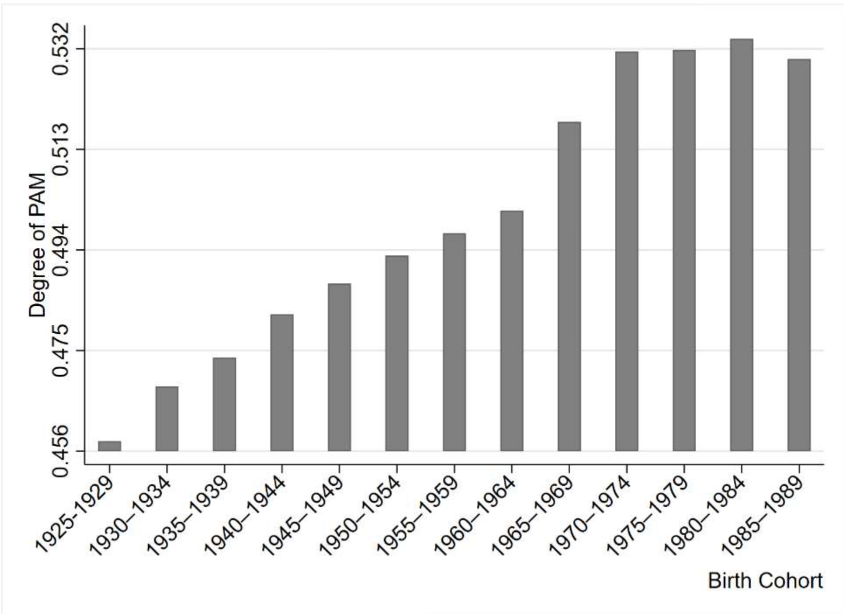
Note: This figure presents educational PAM by five-year birth cohort using the aggregate likelihood ratio measure, based on Euclidean similarity. For comparison, see the cosine similarity-based results in Figure 13.

Figure C.2.3. Trends in Educational PAM by Birth Cohort Using the Weighted Similarity Measure under the Euclidean Similarity Matrix



Note: This figure presents educational PAM by five-year birth cohort using the weighted similarity measure, based on Euclidean similarity. For comparison, see the cosine similarity-based results in Figure 14.

Figure C.2.4. Trends in Educational PAM by Birth Cohort Using the Normalized Trace Measure under the Euclidean Similarity Matrix



Note: This figure presents educational PAM by five-year birth cohort using the normalized trace measure, based on Euclidean similarity. For comparison, see the cosine similarity-based results in Figure 15.

Table C.2.1. Average Educational Euclidean Similarity Scores Across the Full Sample, 2004-2016.

Educational attainment	LHS	HS	SC	C
LHS	0.891	0.675	0.589	0.458
HS	0.698	0.916	0.787	0.570
SC	0.594	0.793	0.902	0.666
C	0.458	0.564	0.638	0.898

Note: This table reports the average pairwise Euclidean similarity scores among educational groups, calculated across all survey years from 2004 to 2016. Similarity values reflect the alignment of educational attributes between men (rows) and women (columns), based on gender-specific vectors derived from MEPS data. Educational categories are abbreviated as follows: LHS = Less than high school; HS = High school graduate; SC = Some college; C = College degree or higher

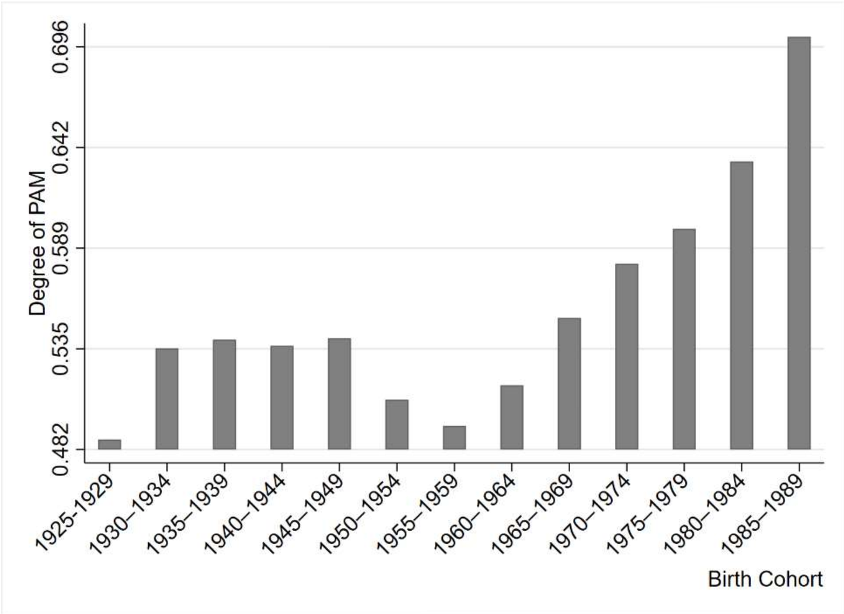
C.3. Results Using the Aggregate Similarity Matrix

In contrast to the occupational and religious analyses, which rely on repeated cross-sectional data, the educational analysis is based on panel data. Consequently, the time-varying similarity matrices for education may capture not only secular changes in educational profiles but also within-cohort, age-related variation—an effect we do not want to include in the analysis.

To address this concern, we conduct a robustness check by constructing a single, time-invariant similarity matrix using pooled observations from all survey years. This aggregate similarity matrix captures the overall structure of educational similarity while minimizing confounding age-related variation.

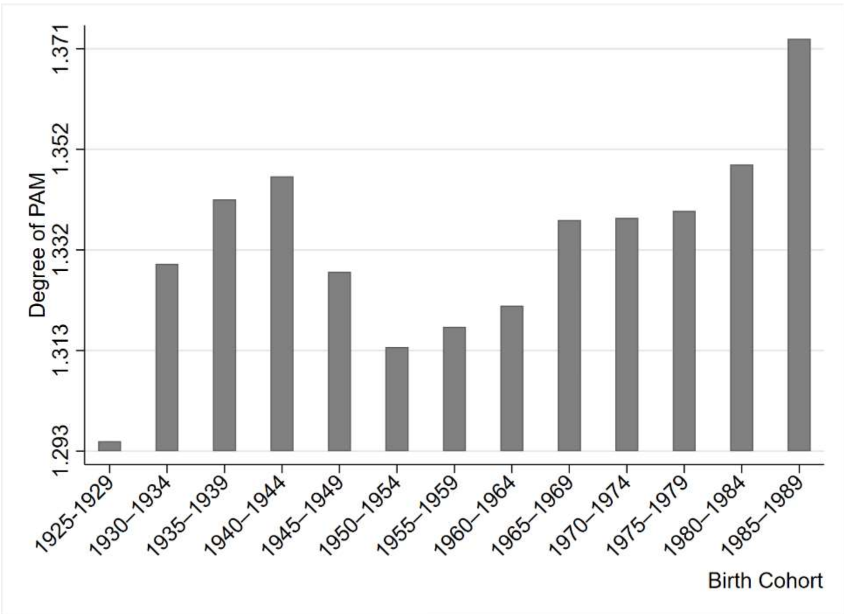
The resulting cohort patterns of educational PAM, based on this time-invariant similarity matrix and shown in Figures C.3.1 through C.3.4, are qualitatively similar to those obtained using year-specific matrices. This consistency suggests that our main findings are not driven by temporal fluctuations in the similarity structure.

Figure C.3.1. Trends in Educational PAM across Birth Cohorts Using the Perfect-Random Normalization Measure under the Aggregate Cosine Similarity Matrix



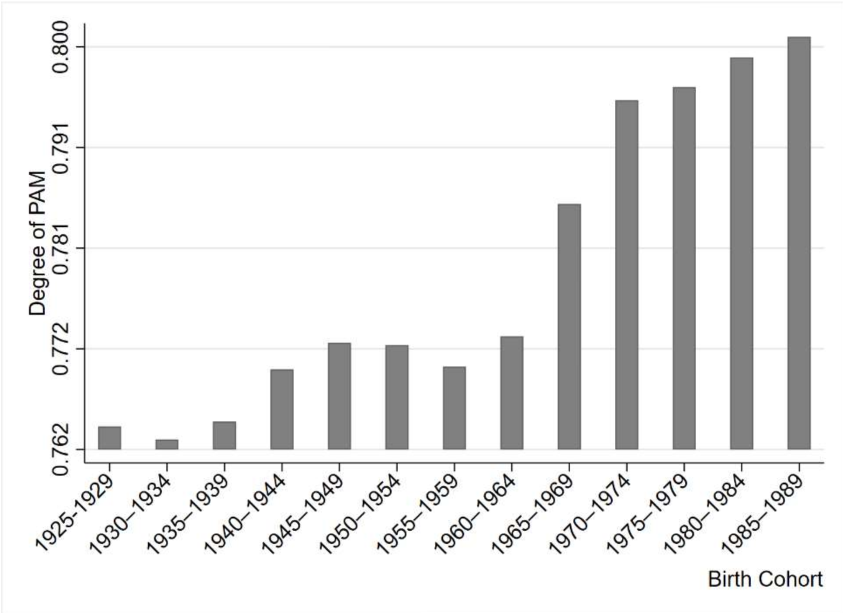
Note: This figure presents educational PAM by 5-year birth cohort using the perfect-random normalization measure. For comparison, see Figure 12 using time-varying cosine similarity matrices.

Figure C.3.2. Trends in Educational PAM across Birth Cohorts Using the Aggregate Likelihood Ratio Measure under the Aggregate Cosine Similarity Matrix



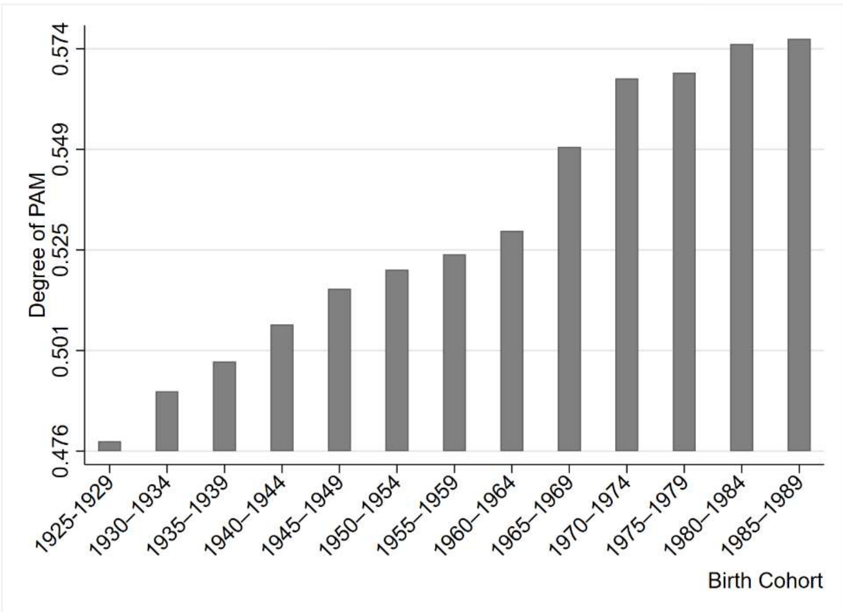
Note: This figure presents educational PAM by 5-year birth cohort using the aggregate likelihood ratio measure. For comparison, see Figure 13 using time-varying cosine similarity matrices.

Figure C.3.3. Trends in Educational PAM across Birth Cohorts Using the Weighted Similarity Measure under the Aggregate Cosine Similarity Matrix



Note: This figure presents educational PAM by 5-year birth cohort using the weighted similarity measure. For comparison, see Figure 14 using time-varying cosine similarity matrices.

Figure C.3.4. Trends in Educational PAM by Birth Cohort Using the Normalized Trace Measure under the Aggregate Cosine Similarity Matrix



Note: This figure presents educational PAM by 5-year birth cohort using the normalized trace measure. For comparison, see Figure 15 using time-varying cosine similarity matrices.

Appendix C.4. List of Educational Attributes

This section provides the full list of 34 attributes used to construct the educational similarity matrices underlying our measures of educational PAM. The attributes are drawn from the MEPS database and include employment status, personal income, and wages or salary, along with thirty health-related indicators encompassing physical and mental health, health-related beliefs, and smoking behavior.

Table C.4.1. List of Educational Attributes

Variable
Wage or Salary ($2009)
Total personal income ($2009)
Employment status
General health status
Smoke cigarettes now (inverse)
Felt everything an effort, past 30 days (adults) (inverse)
How often felt hopeless, past 30 days (adults) (inverse)
How often felt nervous, past 30 days (adults) (inverse)
How often felt restless, past 30 days (adults) (inverse)
How often felt sad, past 30 days (adults) (inverse)
How often felt worthless, past 30 days (adults) (inverse)
Little interest in doing things: past two weeks (inverse)
Feeling down, depressed, or hopeless: past two weeks (inverse)
Blood pressure checked by health professional: last 2 years (inverse)
Health now limits moderate activities (inverse)
Health now limits climbing several flights of stairs (inverse)
Accomplished less because of physical health: past 4 weeks (inverse)
Limited in kind of work because of physical health: past 4 weeks (inverse)
Accomplished less because of emotional problems: past 4 weeks (inverse)
Did work less carefully because of emotional problems: past 4 weeks (inverse)
Pain interfered with normal work: past 4 weeks (inverse)
Blood pressure checked by health professional: last 2 years (inverse)
Health now limits moderate activities (inverse)
Health now limits climbing several flights of stairs (inverse)
Accomplished less because of physical health: past 4 weeks (inverse)
Limited in kind of work because of physical health: past 4 weeks (inverse)
Pain interfered with normal work: past 4 weeks (inverse)
Felt calm/peaceful: past 4 weeks
Had a lot of energy: past 4 weeks
Felt depressed: past 4 weeks (inverse)
Healthy, don't need health insurance

Health insurance not worth cost
More likely to take risks
Can overcome illness without medical help