Alternative Methods to Test for Positive Assortative Mating

Hyunjae Kang
Kyoto University

Steven Stern
Stony Brook University

May 2025

Abstract

We propose hypothesis testing methods to examine positive assortative mating that impose fewer restrictions than existing approaches, such as Siow (2015). Our pseudo-Wald test performs better under weak assortative mating and moderate sample size, leveraging the full information of the sample. The test statistics provide closed-form expressions, making them straightforward to implement. Empirical applications using CPS data confirm that our methods are robust in testing positive assortative mating on relative wage.

1 Introduction

This paper proposes easy hypothesis testing methods to test for positive assortative mating (PAM). A number of empirical studies on sorting in the marriage market have investigated the intensity of assortative mating and how the strength of assortative mating has changed over time using structural (Greenwood et al., 2014; Chiappori et al., 2017; Gualdani and Sinha, 2023) or non-structural (Liu and Lu, 2006; Eika et al., 2019) approaches. Meanwhile, only a handful of econometric tests have been developed for testing PAM that practitioners can readily apply without specifying a structural model (Anderson and Leo, 2013; Siow, 2015).

Broadly, the literature examining PAM features two main approaches. The first constructs indices of sorting, typically by comparing observed match probabilities to counterfactual match probabilities under random pairing. This method measures the intensity of sorting without explicitly modeling individual preferences. Notable examples in this vein include Liu and Lu (2006) and Eika et al. (2019).[1] While Liu and Lu (2006) construct descriptive indices of assortative mating and Eika et al. (2019) develop indices and further examine

---

[1] Fernández and Rogerson (2001), Greenwood et al. (2014), and Abbott et al. (2019) apply similar ideas in the marital matching stage of their structural models.

trends and decompositions, neither study provides a formal hypothesis test or statistical inference procedure to determine the presence of positive assortative mating.

Another strand of literature is a choice-based framework initiated by Choo and Siow (2006b). They propose a marriage matching function, which is based on the transferable utility matching framework with random preferences.[2,3] The critical assumptions of Choo and Siow (2006b) are (a) additive separability of joint surplus, (b) a large marriage market, and (c) an extreme value distribution of the shock (Galichon and Salanié, 2021a). Choo and Siow (2006a) applied the marriage matching function to quantify the effect of the legalization of abortion on gains to marriage.

Building on the transferable utility framework, subsequent studies have extended the Choo and Siow (2006b) model in multiple directions. For instance, Chiappori et al. (2017) incorporate parental investment into a structural marriage model, embedding it in the original Choo and Siow formulation to measure changing returns to education in the U.S. marriage market. Mourifié and Siow (2021) introduce peer effects, Dupuy and Galichon (2014) adapt the framework to continuous attributes using Dagsvik's (1994; 2000) continuous choice structure, and Galichon and Salanié (2021b) show that the extreme value assumption can be relaxed. In a related vein, Fernández and Rogerson (2001) use a search model that delivers PAM but focus on how the degree of sorting correlates with wage gaps and other economic variables. While these structural approaches are powerful for understanding matching behaviors, they often require significant effort in terms of modeling and computation.

Siow (2015) proposes a test based on maximum likelihood estimation (MLE) within the Choo and Siow (2006a) framework, mainly focusing on two key restrictions: total positivity of order two (TP2) and diagonal positivity of order two (DP2). While these constraints can be effective for strong assortative matching, they often perform poorly or become overly restrictive when assortative mating is weak, and is less effective when sample size is small or moderate. Also, the associated test statistics exhibit non-standard asymptotic distributions.

Compared to existing methods for testing PAM, our methods are straightforward to implement and impose no restrictions on the data-generating process. They exploit the full sample information and are robust to outliers. Our tests provide closed-form solutions and thus do not require numerical optimization.

We first define our pseudo-Wald and Likelihood Ratio tests and investigate their distributional properties. The resulting closed-form test statistics are straightforward to implement and computationally efficient, offering strong power while maintaining ease of use. Secondly, we compare our methods with

---

[2]See Chiappori and Salanié (2016) for the survey of the empirical marriage matching models under different assumptions: transferable utility, non-transferable utility, and imperfectly transferable utility models.

[3]See Fox (2010), Fox (2018), and Fox et al. (2018) for an alternative method employing a maximum score approach based on the rank-order property associated with sorting. See Suen and Lui (1999) for a correlation-based test.

the restriction-based method of Siow (2015). Our pseudo-Wald test is more robust in weak assortative matching and performs better when sample size is small ($N = 500$) or moderate ($N = 2,000$). For large sample size ($N = 120,000$), both tests provide support for PAM.

Finally, we provide two empirical applications of matching based on wage and matching in same-sex marriage. Our tests conclude PAM on predicted wage while tests of Siow (2015) find PAM on education. Both tests find PAM in same-sex marriage.

We explain our proposed tests in Section 2. In Section 3, we compare the performance of both approaches under varying intensities of assortative mating. Section 4 provides empirical examples. We conclude in Section 5.

2 Alternative Tests

Our tests differ from Siow (2015) in the way they utilize unrestricted MLE estimates of the matching cells. While Siow (2015) performs a likelihood ratio test with different restrictions, we directly evaluate the test statistics using the estimates of matching cells. In this section, we explain the differences of these two methods.

2.1 Review of Siow (2015)

Suppose there are I types of men, $i = 1, 2, \ldots, I$, and J types of women, $j = 1, 2, \ldots, J$, where types are vertically ordered. Let n_{ij} be the number of observations where a type-i male is matched with a type-j female, and define

$$N = \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij}$$

as the total number of sampled marriages. Let p_{ij} be the probability that a randomly sampled marriage is of type ij. Following Siow (2015), consider the unrestricted likelihood problem:

$$\max_{\{p_{ij}\}} \quad \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij} \log p_{ij} \tag{1}$$

$$\text{subject to} \quad \sum_{i=1}^{I} \sum_{j=1}^{J} p_{ij} = 1.$$

The unrestricted MLE of $\{p_{ij}\}$ is

$$\widehat{p}^u = \{\widehat{p}^u_{ij}\} \quad \text{with} \quad \widehat{p}^u_{ij} = \frac{n_{ij}}{N}.$$

Siow (2015) and related work note that one way to test for PAM is to re-estimate $\widehat{p}$ under restrictions consistent with PAM, and then perform a likelihood ratio test. The main restrictions considered in Siow (2015) are total

3

positivity of order 2 (TP2) and diagonal positivity of order 2 (DP2).[4] Both restrictions rely on the local log-odds ratio for (i,j), which is defined as

$$R_{ij} = \log p_{ij} + \log p_{i+1,j+1} - \log p_{i,j+1} - \log p_{i+1,j}. \qquad (2)$$

Here, R_{ij} measures how strongly types i and j complement each other relative to cross-type matches $(i, j+1)$ and $(i+1, j)$.

TP2 Restriction. The TP2 condition states that *all* such local log-odds ratios are strictly positive for all matching cells:

$$R_{ij} > 0 \quad \text{for all } i = 1, 2, \ldots, I-1 \text{ and } j = 1, 2, \ldots, J-1 \qquad (3)$$

which is equivalent to the case that the underlying matching surplus function is supermodular (Siow, 2015).[5]

DP2 Restriction. DP2 focuses on local log-odds ratios only along the *same type* diagonal (e.g., $(i, i+1)$), which is defined as

$$R_{ii} > 0 \quad \text{for all } i = 1, 2, \ldots, \min(I, J) - 1. \qquad (4)$$

In addition to the primary TP2 and DP2 constraints, Siow (2015) proposes several modifications. In particular, **TP2$'$** imposes the condition $R_{ij} > 0$ except for the top-right and bottom-left matchings. While this exclusion aims to mitigate the influence of extreme or sparsely populated matches, it can arbitrarily discard valuable information; moreover, outliers may arise in other corner cells not explicitly omitted by this rule.

On the one hand, it is not clear which of the restrictions in Siow (2015) are more suitable to assess PAM.[6] For example, the log-likelihood with the TP2 restriction is rejected against the unrestricted likelihood, but the log-likelihood with the DP2 restriction is not rejected. Siow (2015) states that the rejection of the TP2 likelihood is driven by the couples whose completed education is distant. As the extreme matches account only for less than 0.2% of marriages, the TP2 restriction is arguably sensitive to such outliers.[7]

On the other hand, the null hypothesis in Siow (2015) is a space with positive measure, and the alternative hypothesis is a region with positive measure. To see this, consider a simple case where there are just two types of men and two types of women. The parameter space is the unit square, and the condition for PAM is

$$p_{11} + p_{22} - p_{12} - p_{21} \geq 0$$

---

[4]The term "order" refers to the dimension of the determinants for the matching matrix.

[5]See Choo and Siow (2006b) for the descriptions of the underlying structural model.

[6]Chiappori et al. (2017) also propose a test based on their structural model that incorporates returns to parental investment. Their method requires using minimum distance of the moment conditions that come from their structural model. The main purpose of this test is to examine the changes in the intensities of educational homogamy.

[7]Anderson and Leo (2013) develop a test statistic very similar to the TP2 test in Choo and Siow (2006b).

which simplifies to

$$p_{11} + p_{22} \geq 1. \tag{5}$$

The region consistent with the null hypothesis is all values of (p_{11}, p_{22}) that satisfy equation (5), and the region consistent with the alternative hypothesis is the part of the unit square that does not satisfy equation (5). This is an uncommon problem in the literature. Meanwhile, our testing problem is standard: the null hypothesis is a point, and the alternative is a region surrounding the point.

A similar example of interest in the literature is testing a unit root hypothesis for a time series. Almost all of the literature on testing for a unit root assumes that the null hypothesis is nonstationarity. A strong reason for making this assumption instead of stationarity being the null hypothesis is that, when nonstationarity is the null hypothesis, it can be represented by a single $AR(1)$ parameter equalling 1. Thus, the situation is turned into a standard case where the null hypothesis corresponds to a point. However, from an economic point of view, the null hypothesis should be stationarity in that the model behaves better when there is stationarity (we would like to not reject the null hypothesis and then rely on the nice properties of stationary processes). Alternatively, Kwiatkowski et al. (1992), Leybourne and McCabe (1994), and Xiao (2001) all use a null hypothesis of stationarity by constructing a time series model with a random walk where the random walk is the only source of nonstationarity. The null hypothesis is then that the variance of the random walk error is equal to zero.

A second example, much closer to our problem, is to test whether the Slutsky substitution matrix is negative semidefinite. Gill and Lewbel (1992) choose a null hypothesis of semidefiniteness (which has positive measure) and show how to construct a consistent test statistic. Gill and Lewbel (1992) state that the asymptotics of this problem are very similar to that in Kodde and Palm (1986) where the null and alternative hypotheses have positive measure.[8] An easier approach is to construct a Wald test, requiring only the unrestricted estimates.[9] Though it does not deal with the nonstandardardness of the test, it avoids the cost associated with imposing the positive assortative mating restrictions, and it allows for more general tests.

Researchers have developed different methods to test hypotheses related to assortative mating, although not all directly focus on assessing the positivity of assortative mating. For instance, Chiappori et al. (2017) focus on the change of the degree of educational assortative mating over time. They reject the null hypothesis that the degree of educational assortative mating has remained constant over time. Mourifié and Siow (2021) propose a Cobb-Douglas marriage matching function and emphasize the significance of peer and scale effects in marital matching. Gualdani and Sinha (2023) employ a set identification

---

[8]Unfortunately, Cragg and Donald (1996, 1997) show that the asymptotics for a related rank condition are incorrect, and neither paper addresses the correctness of the argument of Gill and Lewbel (1992) about testing for semidefiniteness.

[9]It is not clear which should be considered the restricted model and which the alternative.

approach with nonparametric distributional assumptions on unobserved heterogeneity. Their findings indicate the presence of positive educational sorting among less educated couples.

2.2 A Pseudo-Wald Test

In the next two subsections, we develop a pseudo-Wald test, a likelihood ratio test, and a Lagrange multiplier test. We start with the pseudo-Wald test. Using the local odds-ratios defined in equation (2), we test the null hypothesis that $R_{ij} = 0 \ \forall ij$ with power against the alternative of (positive or negative) assortative mating with

$$\sum_{i,j} \frac{\widehat{R}_{ij}^2}{s_{ij}^2} \sim \chi_K^2$$

where

$$\widehat{R}_{ij} = \left(\log \widehat{p}_{ij}^u + \log \widehat{p}_{i+1,j+1}^u\right) - \left(\log \widehat{p}_{i+1,j}^u + \log \widehat{p}_{i,j+1}^u\right),$$

$$\begin{aligned} s_{ij}^2 &= Var\left[\left(\log \widehat{p}_{ij}^u + \log \widehat{p}_{i+1,j+1}^u\right) - \left(\log \widehat{p}_{i+1,j}^u + \log \widehat{p}_{i,j+1}^u\right)\right] \\ &\approx \frac{(1-p_{ij})}{Np_{ij}} + \frac{(1-p_{i+1j+1})}{Np_{i+1j+1}} + \frac{(1-p_{i+1,j})}{Np_{i+1,j}} + \frac{(1-p_{i,j+1})}{Np_{i,j+1}}, \end{aligned}$$

using a first-order Taylor series approximation, and

$$K = (I - 1)(J - 1).$$

Contrast to Siow (2015) which uses the criteria for PAM (such as TP2 or DP2) as a null, our pseudo-Wald test adopts non-assortative mating as a null hypothesis.

Our pseudo-Wald statistic aggregates all local log-odds ratios $\{\widehat{R}_{ij}\}$ in a single summation which allows us to exploit *all* information in the sample simultaneously without discarding thinly populated cells or imposing stronger global restrictions. When some matching cells are close to zero and others are far from zero, the pseudo-Wald test statistic captures the net effect—i.e., whether the overall "signal" of PAM outweighs zero assortative matching in certain subgroups.

2.3 Likelihood Ratio Test and Lagrange Multiplier Test

Also note that the MLE of p under the restriction of random mating is $\widehat{p}^r = \{\widehat{p}_{ij}^r\}$ where

$$\widehat{p}_{ij}^r = \widehat{x}_i \widehat{y}_j, \quad \widehat{x}_i = \frac{\sum_j n_{ij}}{N}, \quad \widehat{y}_j = \frac{\sum_i n_{ij}}{N}.$$

Thus, one could test the null hypothesis of random mating against a general alternative with a likelihood ratio (LR) test,

$$
\begin{aligned}
T_{LR} &= 2\left[\sum_{i,j} n_{ij}\log\widehat{p}_{ij}^{u} - \sum_{i,j} n_{ij}\log\widehat{p}_{ij}^{r}\right] \\
&= 2\left[\sum_{i,j} n_{ij}\log\frac{\widehat{p}_{ij}^{u}}{\widehat{p}_{ij}^{r}}\right] \\
&= 2\left[\sum_{i,j} n_{ij}\log\frac{\widehat{p}_{ij}^{u}}{\left(\sum_k \widehat{p}_{ik}^{u}\right)\left(\sum_k \widehat{p}_{kj}^{u}\right)}\right] \sim \chi_K^2
\end{aligned}
$$

(which requires only estimation of $\widehat{p}^{u}$) with

$$
K = IJ - 1
$$

which is larger than the degrees of freedom in for the Wald test. This difference exists because, for the Wald test, one loses degrees of freedom at $i = I$ and $j = J$ as it requires to have $i = I + 1$ and $j = J + 1$. This is not the case for the LR test.

Alternatively, one could construct a Lagrange Multiplier (LM) test. Let

$$
L = \sum_{i,j} n_{ij}\log p_{ij}.
$$

Then,

$$
\begin{aligned}
\frac{\partial L}{\partial p_{ij}} \quad \Big|_{\ p_{ij}=\widehat{p}_{ij}^{r}} &= \frac{n_{ij}}{\widehat{p}_{ij}^{r}} \\
&= \frac{n_{ij}}{\frac{\sum_k n_{ik}}{N}\frac{\sum_k n_{kj}}{N}} \\
&= \frac{N\widehat{p}_{ij}^{u}}{\left(\sum_k \widehat{p}_{ik}^{u}\right)\left(\sum_k \widehat{p}_{kj}^{u}\right)}
\end{aligned}
$$

implying that the LM has the same form and the same distribution as the LR test.

3   Properties of the Proposed Tests

The simulation-based finite-sample analysis in this section offers a pragmatic alternative to asymptotic theory, when closed-form test statistic is not available. We first specify an auxiliary model of assortative mating, which we use to (i) simulate the distribution of our test statistics and (ii) to compare our test methods with the tests of Siow (2015). Define a male as having an integer
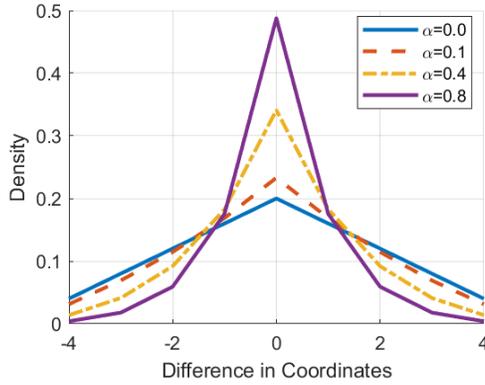
Figure 1: Density from 1 to M (F)

value of x_i between 1 and 5, and define a female as having an integer value of x_j between 1 and 5. Assume that the density of matches in an economy is

$$f(x_i, x_j) = \frac{\exp\{-\alpha\,|x_i - x_j| + \varepsilon_{i,j}\}}{\sum_{x_i'=1}^{I}\sum_{x_j'=1}^{J}\exp\{-\alpha\,|x_i' - x_j'| + \varepsilon_{i,j}\}} \qquad (6)$$

where $\varepsilon_{i,j} \sim N(0, \sigma^2)$.[10,11] We first analyze the distribution of test statistics.[12] Then, we empirically compare performances of our tests with the tests in Siow (2015).

3.1   Distribution of Test Statistics

Figure 1 shows the contour of $f(x_i, x_j)$ as (x_i, x_j) moves from $(1, 5)$ to $(5, 1)$ for $\alpha = 0$, 0.1, 0.4, and 0.8. As α increases, assortative mating becomes stronger, with increased density along the diagonal where difference in coordinates is 0.

Figure 2 shows the distributions of the pseudo-Wald test statistics for samples of 500, 2,000, and 10,000 when $\alpha = 0$ along with the χ_{16}^2 distribution.[13] The χ_{16}^2 distribution exhibits less variance than the pseudo-Wald test statistics for samples of at least 2,000, and it stochastically dominates the pseudo-Wald test statistic distribution for a sample of 500. The 5% critical value for the χ_{16}^2 distribution is off by a meaningful amount relative to the simulated Wald test statistic distributions. The 5% critical value for a test statistic with a χ_{16}^2 distribution is 26.3, while the pseudo-Wald test statistics with sample sizes 500, 2,000, and 10,000 are respectively 32.3, 32.67, and 31.89. The differences among the three occur because the simulated Wald test statistic provides the

---

[10]Note that $\sum_{x_i=1}^{I}\sum_{x_j=1}^{J} f(x_i, x_j) = 1$.

[11]Also, note that this is somewhat like a triangle kernel function with $\alpha = b^{-1}$ where b is the bandwidth (Kokonendji and Zocchi, 2010).

[12]We fix $\varepsilon_{i,j} = 0$ in analyzing the distribution.

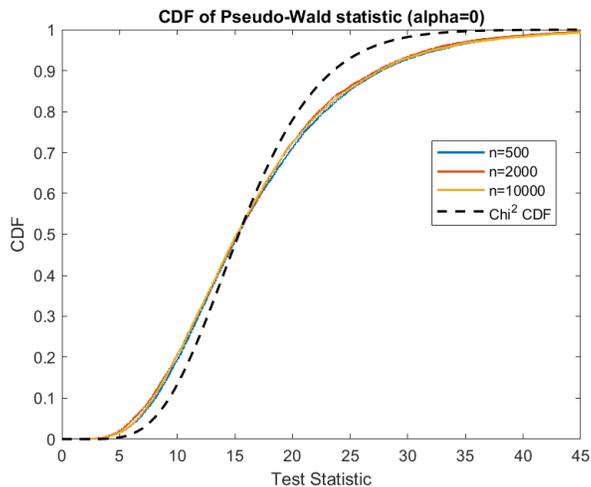[13]The degrees of freedom are $(M-1)(F-1)$.

8

Figure 2: Distribution of Pseudo-Wald Test Statistic for $\alpha = 0$

small-sample distribution of the test statistic, while the simulated LR test relies significantly on asymptotic approximations. These discrepancies arise because the χ^2_{16} curve is an asymptotic approximation, while the simulated distribution reflects the finite-sample distribution of the pseudo-Wald statistic. In practice this mismatch is a minor issue, since researchers can bootstrap the test's null distribution and use those empirical critical values instead of the χ^2_{16}.

Figure 3 shows the distributions of the LR test statistic for samples of 500, 2,000, and 10,000 when $\alpha = 0$ along with the χ^2_{24} distribution. The χ^2_{24} distribution stochastically dominates the LR test statistic distributions for all three sample sizes. The 5% critical value for the χ^2_{24} distribution is off by a meaningful amount relative to the simulated LR test statistic distributions. The 5% critical value for a test statistic with a χ^2_{24} distribution is 36.42, while, for the LR test statistic distributions with sample sizes 500, 1,000, and 10,000, they are respectively 26.54, 26.38, and 26.36.

Figure 4 shows the power functions for both test statistics for different values of α using the 5% critical values under the null hypothesis of $\alpha = 0$. The pseudo-Wald test statistic has weak power for sample sizes less than 10,000, and the LR test statistic has very strong power. Part of the good performance of the LR test statistic relative to the pseudo-Wald test statistic is due to it having more degrees of freedom. Also, the LR test statistic distribution is dominated by the χ^2_{24} distribution (Figure 3) while the pseudo-Wald test statistic is not dominated by the χ^2_{16} distribution (Figure 2); i.e., the LR test statistic has more power than implied by its χ^2_{24} asymptotic distribution while the pseudo-Wald test statistic does not have more power than its χ^2_{16} asymptotic distribution.

9

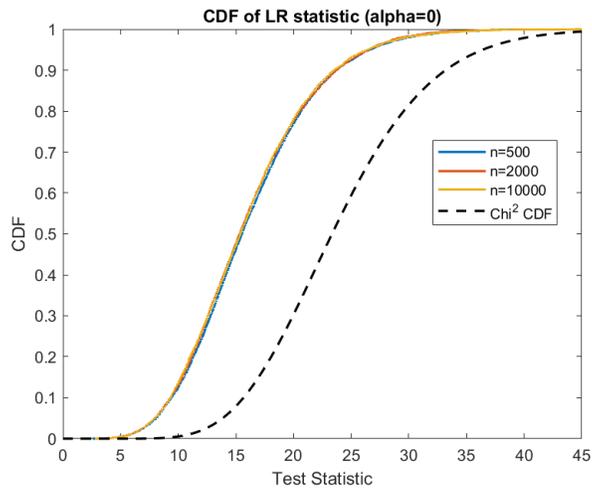Figure 3: Distribution of LR Test Statistic for $\alpha = 0$
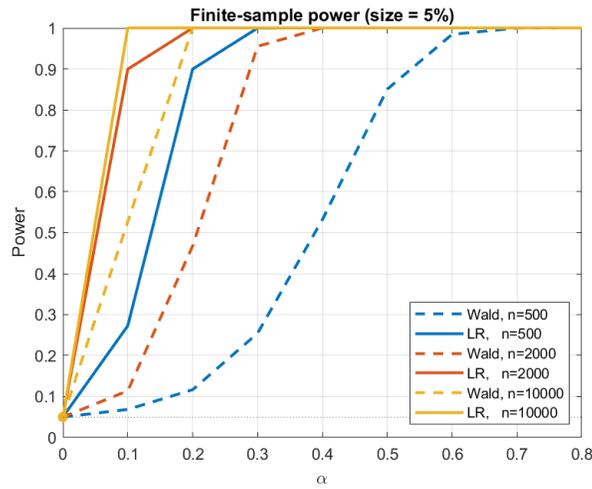

Figure 4: Power for Proposed Tests

3.2 Empirical Comparison

We compare our Wald and Likelihood Ratio (LR) tests with the TP2 and DP2 tests in Siow (2015). The purpose of this exercise is to investigate whether each test can effectively reject PAM when assortative mating is sufficiently weak, and whether each test can detect PAM when assortative mating is present.

The estimated assortative mating intensity α for the 2000 Census dataset is $0.798 \approx 0.8$.[14] Starting from this value, we simulate the DGP and generate synthetic datasets with weaker assortative mating of $\alpha = 0.4, 0.1$ and 0. We use three levels of sample size: 500, 2,000, and 120,000. In this exercise, we fix σ_ε to 0.1. Then, we apply each of the four tests and compare their performance in testing PAM.

Table 1: Verdict for Positive Assortative Mating (PAM) by Four Methods

	$\alpha = 0$				$\alpha = 0.1$				$\alpha = 0.4$				$\alpha = 0.8$			
n	Wald	LR	TP2	DP2	Wald	LR	TP2	DP2	Wald	LR	TP2	DP2	Wald	LR	TP2	DP2
500	N	N	P	P	N	N	P	P	P	P	P	P	P	P	P	P
2000	N	N	P	P	N	N	P	P	P	P	P	P	P	P	P	P
120000	P	P	N	P	P	P	N	P	P	P	N	P	P	P	N	P

Notes: Wald and LR refer to the proposed Pseudo-Wald and Likelihood Ratio Tests, respectively. TP2 and DP2 refer to tests by Siow (2015). "P" indicates PAM, and "N" indicates no PAM.

Table 1 presents the results of applying our Wald and LR tests and the TP2 and DP2 of Siow (2015). N indicates no support for PAM, and P indicates the support for PAM. Since rejection of our tests versus Siow's tests implies opposite conclusions about PAM, this notation is introduced to alleviate confusion. Table 2 reports the detailed results. It is worth noting that $(\alpha, n) = (0.8, 120,000)$ closely matches the application of Siow (2015). Our findings replicate theirs: TP2 is rejected while DP2 is not rejected under the parameter/sample-size pair.

For weak assortative mating of $\alpha = 0$ and 0.1, the two tests yield opposite results. Our proposed pseudo-Wald and LR tests reject PAM for sample sizes of 500 and $2,000$. On the other hand, TP2 and DP2 of Siow (2015) support PAM. The results suggest that under moderate sample sizes, our tests successfully reject PAM when assortative mating intensity is weak ($\alpha = 0.1$) or there is no PAM ($\alpha = 0$). For large sample sizes, our Wald, LR tests and DP2 of Siow (2015) supports PAM.

All tests consistently support PAM when $\alpha = 0.4$ and $\alpha = 0.8$. This is consistent with findings of previous papers analyzing PAM around 2,000 (Greenwood et al., 2014; Chiappori et al., 2017). At the same time, when the sample size is very large, all four tests produce consistent results. Our Wald and LR tests reject the zero assortative mating. For Siow (2015), DP2, as a criteria for PAM, is not rejected, thus support PAM. TP2 is rejected, but again, this is consistent with their results.

---

[14]The same data is used in the empirical application of Siow (2015).

Table 2: Test Statistics by N and α

N	$\alpha = 0.0$	$\alpha = 0.1$	$\alpha = 0.4$	$\alpha = 0.8$
Panel A: Proposed Pseudo-Wald				
500	19.92	20.50	38.12**	96.01***
2000	29.72	35.73*	101.37***	348.36***
120000	847.95***	1093.76***	5183.76***	19926.79***
Panel B: Proposed LR Test				
500	19.23	25.99	110.31***	310.02***
2000	28.71	52.00**	382.76***	1219.64***
120000	782.29***	2225.41***	21820.48***	71478.45***
Panel C: TP2 of Siow (2015)				
500	15.69	11.21	9.17	8.63
2000	23.33	16.76	11.94	10.98
120000	638.38***	401.57***	333.50***	238.77***
Panel D: DP2 of Siow (2015)				
500	2.07	1.24	0.02	$7.8e{-}7$
2000	3.55	1.06	$9.8e{-}5$	$1.2e{-}6$
120000	92.21*	20.35	$1.3e{-}5$	$3.4e{-}5$

Note: * $p \leq 0.10$, ** $p \leq 0.05$, *** $p \leq 0.01$. In computing p, we use χ^2_k for Pseudo-Wald and LR Tests and parametric bootstrap for TP2 and DP2.

Table 3: CPS 2000 Matching Table

(a) Matching Table by Education

	Below High	High	Some Col	Bachelor	Graduate
Below High	0.101	0.048	0.018	0.003	0.001
High	0.040	0.191	0.075	0.021	0.004
Some Col	0.014	0.079	0.118	0.032	0.007
Bachelor	0.002	0.032	0.048	0.073	0.014
Graduate	0.001	0.008	0.016	0.033	0.025

(b) Matching Table by Predicted Wage Quintiles

	Q1	Q2	Q3	Q4	Q5
Q1	0.101	0.050	0.034	0.011	0.004
Q2	0.054	0.040	0.060	0.030	0.016
Q3	0.028	0.042	0.047	0.056	0.027
Q4	0.013	0.039	0.030	0.058	0.060
Q5	0.005	0.029	0.029	0.044	0.093

Notes: Entries are matching probabilities. Rows denote male spouse's category, and columns denote female spouse's category.

4 Empirical Applications

4.1 Assortative Mating on Predicted Wage

We test PAM using predicted wages based on within-gender wage deciles from the CPS March 2000 sample ($N = 12{,}497$ married households). Compared with discrete measures, using wage provides a finer, continuous metric of assortative mating. Also, the rank measure is scale-invariant and robust to outliers.[15] We employ a two-step estimation procedure (Heckman, 1979) and use predicted log wages to construct our predicted wage measures.[16]

Table 5: Comparison of Predicted Wage Quintile Example

Test	Statistic	p-value	Assortative Mating
Kang–Stern (pseudo-Wald)	648.038	< 0.001	P
Kang–Stern (LR)	4420.734	< 0.001	P
Siow TP2	67.031	< 0.001	N
Siow DP2	13.101	< 0.001	N

Note: For proposed methods, the null hypothesis is zero assortative mating. For Siow's methods, the null hypothesis implies PAM. In computing p, we use χ^2_k for Pseudo-Wald ($k = 16$) and LR Tests ($k = 24$) and parametric bootstrap for TP2 and DP2.

Table 3 presents the contingency tables by (a) education and (b) predicted wage quintiles.[17] Compared to the five-category education matrix, the wage-rank table exhibits greater dispersion from the main diagonal, reflecting within-education variation captured by rank measures. The diagonal components in the education matching table (Table 4a) sum to 0.508, but those in the wage-rank matching table (4b) sum to only 0.339, indicating greater dispersion in the predicted wage matching table.

Table 5 reports the results of the four tests using predicted wage deciles. The pseudo-Wald and likelihood-ratio (LR) statistics strongly reject the null of no assortative mating ($p< 0.001$), consistent with the strong PAM found in the literature (Greenwood et al., 2014; Chiappori et al., 2017). Siow's TP2 and DP2 statistics reject PAM, presumably because its restriction based methods are sensitive to the off-diagonal elements.

4.2 Assortative Mating in Same-Sex Marriage

Finally, we analyze same-sex married couples using the 2019 Census data. Table 6 presents the results. The sample size for same-sex married couples is 1,645,

---

[15]Several studies have employed income rank as the measure of assortative mating (Fagereng et al., 2022; Holmlund, 2022; Almås et al., 2023).

[16]We explain the procedure in Appendix B. See Carroll et al. (2021) for the more detailed explanations and the results of the estimation.

[17]While the literature often employs deciles (Fagereng et al., 2022), we use quintiles for fair comparison with educational assortative mating, which usually has five discrete measures. The results of our tests are robust to the number of matching indices.

consisting of 695 male couples and 950 female couples. Our test supports PAM in aggregated, male, and female samples. Both pseudo-Wald and LR tests strongly reject the null of zero assortative mating. Interestingly, Siow's DP2 results indicate PAM for male same-sex couples.[18]

Table 6: Positive Assortative Mating for Same-Sex Married Couples

Sample	Test	Statistic	p-value	Assortative Mating
Both	pseudo-Wald	218.38	< 0.001	P
Both	LR	676.64	< 0.001	P
Both	TP2	3.62	0.507	P
Both	DP2	0	0.447	P
Male	pseudo-Wald	132.23	< 0.001	P
Male	LR	286.52	< 0.001	P
Male	TP2	14.39	0.018	N
Male	DP2	$4.2e{-}6$	0.244	P
Female	pseudo-Wald	103.73	< 0.001	P
Female	LR	410.13	< 0.001	P
Female	TP2	1.64	0.611	P
Female	DP2	$1.7e{-}5$	0.018	N

Note: For proposed methods, the null hypothesis is zero assortative mating. For Siow's methods, the null hypothesis implies PAM. In computing p, we use χ^2_k for Pseudo-Wald and LR Tests and parametric bootstrap for TP2 and DP2. For DP2 results of both samples, the statistic is less than the machine precision, so we report as 0.

5 Conclusion

We develop empirical tests for PAM that are straightforward to implement and outperform existing methods. Our pseudo-Wald test performs well, particularly when assortative mating is weaker and sample sizes are moderate, providing advantages over existing tests. Empirical applications using CPS and Census data confirm that our methods effectively detect PAM on wage, confirming their practical value.

---

[18]In contrast, Ciscato et al. (2020) find weaker assortative mating for male same-sex couples using California samples between 2008-2012.

References

Abbott, B., Gallipoli, G., Meghir, C., and Violante, G. L. (2019). Education policy and intergenerational transfers in equilibrium. *Journal of Political Economy*, 127(6):2569–2624.

Almås, I., Kotsadam, A., Moen, E. R., and Røed, K. (2023). The economics of hypergamy. *Journal of Human Resources*, 58(1):260–281.

Anderson, G. and Leo, T. W. (2013). An empirical examination of matching theories: The one child policy, partner choice and matching intensity in urban china. *Journal of Comparative Economics*, 41(2):468–489.

Carroll, M., Kang, H., and Stern, S. (2021). The relationship between spouses' wages over time. Unpublished manuscript.

Chiappori, P.-A. and Salanié, B. (2016). The econometrics of matching models. *Journal of Economic Literature*, 54(3):832–861.

Chiappori, P.-A., Salanié, B., and Weiss, Y. (2017). Partner choice, investment in children, and the marital college premium. *American Economic Review*, 107(8):2109–2167.

Choo, E. and Siow, A. (2006a). Estimating a marriage matching model with spillover effects. *Demography*, 43(3):463–490.

Choo, E. and Siow, A. (2006b). Who marries whom and why. *Journal of Political Economy*, 114(1):175–201.

Ciscato, E., Galichon, A., and Goussé, M. (2020). Like attract like? a structural comparison of homogamy across same-sex and different-sex households. *Journal of Political Economy*, 128(2):740–781.

Dagsvik, J. K. (1994). Discrete and continuous choice, max-stable processes, and independence from irrelevant attributes. *Econometrica*, 62(5):1179–1205.

Dagsvik, J. K. (2000). Aggregation in matching markets. *International Economic Review*, 41(1):27–58.

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

Eika, L., Mogstad, M., and Zafar, B. (2019). Educational assortative mating and household income inequality. *Journal of Political Economy*, 127(6):2795–2835.

Fagereng, A., Guiso, L., and Pistaferri, L. (2022). Assortative mating and wealth inequality.

Fernández, R. and Rogerson, R. (2001). Sorting and long-run inequality. *The Quarterly Journal of Economics*, 116(4):1305–1341.

Fox, J. (2010). Estimating the employer switching costs and wage responses of forward-looking engineers. *Journal of Labor Economics*, 28(2):357–412.

Fox, J. (2018). Estimating matching games with transfers. *Quantitative Economics*, 9:1–38.

Fox, J., Yang, C., and Hsu, D. H. (2018). Unobserved heterogeneity in matching games. *Journal of Political Economy*, 126(4):1339–1373.

Galichon, A. and Salanié, B. (2021a). Cupid's invisible hand: Social surplus and identification in matching models. *arXiv preprint arXiv:2106.02371*.

Galichon, A. and Salanié, B. (2021b). Cupid's invisible hand: Social surplus and identification in matching models. *Review of Economic Studies*. Forthcoming.

Greenwood, J., Guner, N., Kocharkov, G., and Santos, C. (2014). Marry your like: Assortative mating and income inequality. *American Economic Review*, 104(5):348–353.

Gualdani, C. and Sinha, S. (2023). Partial identification in matching models for the marriage market. *Journal of Political Economy*. Forthcoming.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1):153–161.

Holmlund, H. (2022). How much does marital sorting contribute to intergenerational socioeconomic persistence? *Journal of Human Resources*, 57(2):372–399.

Kokonendji, C. and Zocchi, S. (2010). Extensions of discrete triangular distributions and boundary bias in kernel estimation for discrete functions. *Statistics & Probability Letters*, 80(21–22):1655–1662.

Liu, H. and Lu, J. (2006). Measuring the degree of assortative mating. *Economics Letters*, 92(3):317–322.

Mourifié, I. and Siow, A. (2021). The Cobb-Douglas Marriage Matching Function: Marriage Matching with Peer and Scale Effects. *Journal of Labor Economics*, 39(S1):S239–S274.

Siow, A. (2015). Testing becker's theory of positive assortative matching. *Journal of Labor Economics*, 33(2):409–441.

Suen, W. and Lui, H.-K. (1999). A direct test of the efficient marriage market hypothesis. *Economic Inquiry*, 37(1):29–46.

Appendix

A   Data

2019 Current Population Survey (CPS) ASEC

We use 2000 and 2019 CPS for the wage and the same-sex marriage analysis, respectively. The working sample is defined by the following sequential selection rules:

1. **Age.** Keep only households in which both spouses are aged 16–64. This removes roughly 7–13 % of observations, depending on year

2. **Marital status.** Retain legally married couples; exclude widowed, divorced, never-married and co-habiting pairs This step eliminates about one-half of the raw sample in recent years.

3. **Employment class.** Exclude households in which either spouse is self-employed, works for government, or is on active military duty, following Siow (2015). The rule removes 10–15 % of couples each year

4. **Missing or implausible data.** Drop observations with missing schooling, race, region, or hours/weeks information and discard hourly wages below the year-specific real federal minimum wage

2000 Census

We replicate the selection rules used by Siow (2015). Because identical filters are applied, our Census contingency tables are identical to those in Siow (2015).

B   Heckman Two-Step Wage Prediction

First, for each gender and year we estimate a probit

$$y_{ij}^* = X_{ij}\beta + \varepsilon_{ij}, \qquad y_{ij} = 1\{y_{ij}^* > 0\},$$

where X_{ij} includes a quadratic in age, race dummies, four education dummies, region, metro status and, for women, an indicator of at least one child under 5 years old.

Second, on the working subsample we estimate

$$w_{ij} = Z_{ij}\gamma + \sigma_u \rho\, \lambda_{ij} + u_{ij},$$

where $Z_{ij} = X_{ij}$ excluding the young-child variable (an exclusion restriction) and $\lambda_{ij} = \varphi(X_{ij}\hat{\beta})/\Phi(X_{ij}\hat{\beta})$ is the inverse Mills ratio. Then, We compute

$$\tilde{w}_{ij} = \begin{cases} Z_{ij}\hat{\gamma} + \hat{\sigma}_u\hat{\rho}\,\lambda_{ij}, & \text{if } y_{ij} = 1, \\ Z_{ij}\hat{\gamma} - \hat{\sigma}_u\hat{\rho}\,\lambda_{ij}, & \text{if } y_{ij} = 0, \end{cases}$$

and use $\tilde{w}_{ij}$ to generate quintiles of wage.