Technical Paper Series
Congressional Budget Office
Washington, DC

MATE MATCHING FOR MICROSIMULATION MODELS

Kevin Perese
(Kevin.Perese@cbo.gov)
Long-Term Modeling Group
Congressional Budget Office
Washington, DC

November 2002
2002-3

 The author would like to thank Amy Harris, Josh O'Harra, John Sabelhaus, Michael Simpson, and Joel Smith of CBO for their support and helpful comments.

Abstract

In dynamic microsimulation models, the process of determining who marries whom in the annual marriage market has important implications for numerous other processes modeled. There are two broad methodologies used in most microsimulation models to match individuals with each other – a stable marriage algorithm approach or a stochastic framework approach. This paper reviews those two approaches and summarizes the mate-matching techniques used in other dynamic microsimulation models. Building upon the methods used in other models, this paper presents an extension to the stochastic mate-matching approach. This new technique accurately replicates the joint distribution of spousal characteristics (differences in age, education, and earnings) observed in survey data. This technique is used in the Congressional Budget Office's Long-Term (CBOLT) microsimulation model.

1. Introduction

Dynamic microsimulation is a modeling technique that incrementally ages a population while projecting individual demographic and economic events such as marriage, fertility, employment, savings, divorce, retirement, and finally, mortality. Because dynamic microsimulation models produce longitudinal projections that contain detailed micro-level demographic and employment histories, they are excellent tools for analyzing the distribution of future retirement income and for estimating the impact of prospective Social Security reforms. A comprehensive dynamic microsimulation model consists of many modules that are responsible for projecting each of the included economic and demographic processes.

The demographic process of marriage is an integral part of dynamic microsimulation models. In most models, the construction of synthetic marriages is a two-step process. The first step involves estimating individual probabilities of entering into marriage in a given year (assuming the individual was not married in the previous year). Based on Monte Carlo techniques, those marriage probabilities are compared with uniformly distributed random numbers, and men and women are sent to an annual "marriage market" to be matched with each other.[1] The second step applies additional statistical techniques and computational algorithms to determine who will be matched with whom. This paper focuses on the latter of those steps – the processing of annual marriage markets to produce synthetic marriages that resemble real-world marriages. The detailed description of mate-

1 See O'Harra and Sabelhaus (2002) for a detailed description of an approach to modeling marital transitions in a long-term dynamic microsimulation model.

matching techniques presented here will serve as the foundation for future papers that will investigate the policy implications of simulated joint-spousal characteristics.

The annual process of matching single individuals to form new, synthetic families is an important facet of the many interactions that microsimulation models attempt to capture. Because dynamic microsimulation models rely on strong internal dependencies, inaccuracies in one module can have ripple effects throughout the model. The quality of matches produced is one such process that can have important ramifications on broad model results. For example, the quality of the marital matches produced will affect the accuracy of individual-level Social Security eligibility and benefit amounts. Age and income differences in married couples directly affect whether benefits will be granted on the basis of one's own earnings record or the earnings record of a spouse. In addition, the collection and duration of spouse and survivors benefits are closely tied to spousal age differences. Other key economic behaviors (such as employment, wages, and savings) and primary demographic processes (such as fertility and mortality) are also dependent on the marital status and joint characteristics of spouses.

The mate-matching problem is succinctly stated as the creation of an efficient algorithm to convert annual sets of potential spouses into synthetic marriages such that matches have appropriate levels of homogamy, or the degree to which likes match with likes, based on selected spousal characteristics. Mate matching attempts to accurately replicate the central tendency as well as the dispersion of the assortative mating patterns observed in the real world. This paper presents alternative mate-matching methodologies available to modelers

and surveys the mate-matching algorithms used in other microsimulation models.[2] Finally, a detailed description of an innovative stochastic methodology is presented along with results from an example simulation using this new technique. The mate-matching technique presented here was developed as part of the microsimulation component of the Congressional Budget Office's Long-Term (CBOLT) model – a complex model that uses macroeconomic simulations, actuarial projections, and microsimulation techniques to simulate Social Security outcomes over a 75-year projection period (see Harris, O'Harra, Page, Perese, Sabelhaus, Simpson, and Smith, 2002).

2. Techniques and Applications

There are two distinct methodologies that have been used to match couples in microsimulation models: a stable marriage approach and a stochastic approach. Both techniques start by measuring the likelihood that a given man would marry a given woman. The two techniques differ, however, by how this information is used to assign individuals to each other as spouses. Whereas the stable marriage approach processes this information using an optimization technique, the stochastic approach processes this information on the basis of Monte Carlo techniques to unite couples.

Several dynamic microsimulation models have been developed over the past couple of decades. Those modeling efforts were heavily drawn upon when designing the mate-matching methodology presented here. Dynamic microsimulation modeling originated at

2 Note that the difference between dynamic microsimulation models and static microsimulation models is simply the inclusion of new cohorts in dynamic models. Because this paper does not focus on the dynamic portion of microsimulation models, "microsimulation" and "dynamic microsimulation" will be used

the Urban Institute in the 1970s when Guy Orcutt, Steven Caldwell, Richard Wertheimer, and Sheila Zedlewski built the DYNASIM model (Orcutt, Caldwell, and Wertheimer, 1976; Zedlewski, 1990). DYNASIM has since gone through two major updates and revisions – first in the mid-1980s and then again in 2000. Several other dynamic microsimulation models have been built since then. Steven Caldwell started a modeling group, Strategic Forecasting, at Cornell University, which has built the CORSIM model. Another notable microsimulation model is DYNACAN. DYNACAN is a model used for Canadian retirement policy analysis and is a direct descendant of the CORSIM model.

CORSIM and DYNACAN have, until recently, relied on a modified stable marriage algorithm to match spouses. After some investigation, however, it was determined that the stable marriage algorithm produced marriages that did not have desirable joint-spousal characteristics. That limitation prompted CORSIM and DYNACAN to adopt a stochastic matching approach. DYNASIM, on the other hand, has always relied on a stochastic matching algorithm. The mate-matching approach discussed in this paper builds on the stochastic algorithms used by CORSIM, DYNACAN, and DYNASIM. Before describing this new technique, this paper presents background information on the stable marriage algorithm and the mate-matching methodologies used in the other microsimulation models.

The Stable Marriage Algorithm:

The stable marriage algorithm is essentially a computer science problem that has existed for a very long time. A paper by Gale and Shapely (1962) defines and solves the stable

---

interchangeably.

marriage problem formally, although an algorithm to match annual pools of residents with hospitals had been in use for more than a decade before the academic publication of the computational algorithm. Several other articles and books have been published since then documenting the algorithm, its computational efficiency, and various permutations of the stable marriage problem and algorithms to handle those permutations (Knuth, 1976; Gusfield and Irving, 1989).

Understanding the stable marriage algorithm hinges on the definition of "stable." A stable set of marriages exists if there are no two couples where a partner in *each couple* would prefer to be matched with a person in the *opposite couple* rather than the person to whom they are matched. For this condition to be satisfied, however, what it means to "prefer" one potential mate over another needs to be defined and quantified. Given a specification of characteristics that spouses desire in each other, for each male_i and female_j a measure of compatibility, denoted c_{ij}, could be calculated. With those measures of compatibility calculated, a stable set of marriages is defined as a state in which for all couples (m_i, f_i) and (m_j, f_j):

$$c_{ii} > c_{ij} \quad \text{or} \quad c_{jj} > c_{ij} \tag{1}$$

Note that only one of those inequalities must be true for the set of matches to be "stable." This means that just one of the matches must be better off than a match with opposite partners. Or, in other words, only if both of these inequalities fail would the set of marriages be "unstable" (Johnson, 2000).

The process of achieving a stable set of marriages is as follows: a measure of compatibility for all potential pairs is calculated and sorted in descending order. The best match is married. Next, all pairings that included either the husband or wife from this marriage are removed from the remaining potential pairs. Then, the compatibility of the remaining potential pairs is re-ranked and the next most compatible couple is married. This process continues recursively until all the matches have been made.

Although the stable marriage algorithm will efficiently match potential spouses with each other, there are several shortcomings to this methodology for microsimulation purposes (Bouffard, Easther, Johnson, Morrison, and Vink, 2001; Easther and Vink, 2000). The most important shortcoming of the stable marriage algorithm lies in the joint characteristics of the spouse matches it creates. Bouffard and others (2001) examined the distributions of differences in age and differences in earnings produced in CORSIM/DYNACAN and compared those against distributions found in census data. They found that the stable marriage algorithm produced annual sets of marriages that have an exorbitant proportion of marriages occurring where the husband is one year older than the wife (relative to census data). Somewhat counterintuitively, the stable marriage algorithm also produced too many "extreme" marriages – or marriages where the difference is spouses' ages is greater than 20 years. This bimodal distribution of marriages is produced with the stable marriage algorithm because after the "best" marriages are made, there are only relatively "bad" matches left to be made. Finally, Bouffard and others

(2001) document a much higher correlation between husband and wife's earnings, relative to census data, produced when using the stable marriage algorithm.[3]

Easther and Vink (2000) show that there are theoretical shortcomings of the stable marriage algorithm that are at the root of those unsatisfactory results. They argue that the stable marriage algorithm misuses the information contained in the "compatibility" measure estimated for each potential pairing. The compatibility measure represents the probability of a union given a specific set of spousal characteristics. The stable marriage algorithm, however, is designed to make "optimal" matches. That goal essentially increases the likelihood of good matches and decreases the likelihood of poor matches. By giving preference to optimal matches, the stable marriage algorithm produces significantly different distributions of matches than the one produced by the probabilistic distribution of the compatibility index. Furthermore, the reliance on fixed preferences to rank all potential matches in the stable marriage algorithm results in a deterministic matching process, which is uncharacteristic of almost all other microsimulation processes.

The Stochastic Approach:

An alternative to the stable marriage algorithm is to use a stochastic approach to matching potential spouses in the marriage market each year. As mentioned previously, both the stable marriage algorithm and a stochastic matching algorithm start with a calculation of the likelihood of union formation between two potential spouses. Rather than relying on an optimization routine to match spouses, a stochastic matching routine

3 The authors indicate that some of the excessive correlation in earnings produced by the stable marriage

processes this information on the basis of Monte Carlo techniques. If the probability of union formation derived in the first step exceeds a random number drawn from a uniform distribution, a marriage occurs.

DYNASIM Methodology:

After estimated marriage equations determine who is to be married in a given year, DYNASIM stochastically matches newlyweds. Unlike the stochastic methodology employed by CORSIM and DYNACAN, DYNASIM's probability of union formation is estimated as part of the actual matching algorithm rather than as a separate process.

DYNASIM's matching algorithm begins by randomly sorting the queue of bachelors and bachelorettes. For the first available bachelor, the probability of marrying the first available bachelorette is calculated with the following distance function:

$$P(union_{mf}) = e^{-0.5\sqrt{[(Age_m - Age_f)^2 + (Edu_m - Edu_f)^2]}} \qquad (2)$$

If this probability exceeds a randomly drawn number, then the match is made. Otherwise, a match is attempted between the first bachelor and the second available bachelorette. These comparisons continue until either a match is made or a match has been attempted with the first 10 available bachelorettes.[4] If no match has been made after 10 attempts, then the match that has the highest probability is married. Individuals who are not

---

algorithm may be attributable to discrepancies in the earnings measures used.

4 The number of females that a male has to attempt a match with is equal to 20 for men age 35 or older, due to their having a potentially larger pool of women to choose from relative to their younger counterparts.

matched (due to excess men or women in the marriage queues) are returned to the unmarried population and are at risk of marriage again in the following year.[5]

CORSIM/DYNACAN Methodology:

As mentioned above, CORSIM and DYNACAN abandoned the stable marriage algorithm in favor of a stochastic framework.[6] The mate-matching algorithm in CORSIM and DYNACAN begins with an estimation of the relative "compatibility" for each potential couple. The compatibility index is estimated using a logistic regression on a potential pairs data file of recent marriages observed in census data.[7] A potential pairs data file consists of a separate observation for every potential match. Each observation contains the characteristics of the potential husband and the characteristics of the potential wife. As such, given a set of *n* newlyweds, a potential pairs data file will contain *n*-squared observations.

The dependent variable in the data file is set equal to one if the observation is an actual marriage. For all other potential matches, the dependent variable is set equal to zero. The set of variables used to determine the compatibility of couples includes difference in age, difference in age squared, difference in years of education, number of children the woman has, race, labor force participation, and difference in earnings. In addition, there are several interaction effects included, such as male education if he is older than the female

5 Note that DYNASIM does not allow interracial marriages to occur, and thus this process is carried out separately for queues of black and white males and females.
6 Note that the mate-matching methods that CORSIM and DYNACAN use are still evolving and that the information presented here represents the most up-to-date information available.
7 Bouffard and others (2001) use the DYNACAN model to explore differences between the stable marriage

and the product of female education and male earnings. In this potential pairs framework, the predicted covariates in the regression are then used to obtain the relative likelihood of any given match (Bouffard and others, 2001).

Unlike DYNASIM, the CORSIM/DYNACAN procedure calculates the probability of a match for every potential couple up front. Those predicted probabilities are summed to create an *n*-by-*n* matrix with cumulative probabilities. Each cell in the matrix represents a potential couple that is appropriately weighted by the predicted probability. A random number between zero and the sum of the predicted probabilities is selected; the cell that number falls within is the couple that gets married. This ensures that a match is made with each random number drawn. All the cells along the row and column of the selected cell are zeroed out to remove the potential pairs that included each member of the newly matched couple. The random selection of couples continues recursively until all matches are made.

Although this technique produces a more favorable distribution of spousal age differences, it has little effect on the incidence of marriages with extreme age differences. In addition, there are only slight increases in the earnings correlation produced by the stochastic method relative to the correlation simulated under the stable marriage algorithm. Those improvements over the stable marriage algorithm are noteworthy, yet due to the shortcomings that remain, further improvements on the matching algorithm used in these two models are being sought (Bouffard and others, 2001).

---

algorithm and a stochastic algorithm. As such, the census data used come from the 1981 Canadian census.

3. An Alternative Stochastic Methodology

The shortcomings of the stable marriage algorithm for use in microsimulation and lack-luster results of extant stochastic techniques provide motivation for extensions and improvements to the methodologies employed by CORSIM, DYNACAN, and DYNASIM. As described in the section on stochastic mate-matching methodologies, the first step involves estimating the probability of a match actually occurring for a given set of characteristics among potential husbands and wives.

Data:

To estimate the likelihood of a union between potential pairs, two logistic regressions are estimated – one for men's first marriages and another for men's higher order marriages. To estimate those models, a family-level data file that contains both husband and wife characteristics is constructed. For each man in the data file, the characteristics of the current real wife are compared with the characteristics of the *n*-1 other women in the data file of newlyweds. Each comparison is output as a separate record. This process creates a potential pairs analysis file with *n*-squared observations. For every observation where the characteristics of a husband's wife matches those of another newlywed wife, the dependent variable is set equal to one; otherwise, it is set equal to zero. This convention produces a likelihood that bachelor_i would choose bachelorette_j if all he was looking for were the set of characteristics he found in his real wife.

Age, education, average lifetime earnings quintile, and marriage number are the characteristics used to define what a man seeks in a wife. So assume a man entering his

first marriage selects a 30-year-old woman who is also in her first marriage, has 16 years of education, and is in the third earnings quintile (relative to other women in the marriage pool). For every observation in the potential pairs data file where this particular man is matched to a woman with those same characteristics (including his actual wife), the dependent variable is set equal to one. This procedure increases the number of marriage events observed in the dependent variable to be greater than *n* in an *n*-squared data file.[8] As this process is performed separately for men's first marriages and men's higher order marriages, two data files are created and separate models are estimated using each file.

Those models are estimated using data from the marriage history topical module in the 1996 Survey of Income and Program Participation (SIPP). In addition, this topical module is linked to longitudinal Social Security earnings records, which is the basis for the calculation of average lifetime earnings quintiles. Those data also provide baseline descriptions of the joint distributions that serve as the benchmark with which simulated outcomes are compared. Based on those data, there are approximately 3,450 newlyweds (or 1,725 couples) available for the analysis. Newlywed couples are defined as marriages that started in 1994, 1995, or 1996. Due to missing information in the analysis variables, approximately 500 couples are dropped from the analysis. The final sample file includes approximately 1,277 couples – 834 first marriage couples and 443 higher order marriage couples.[9]

8 Among first marriages, there is a six-fold increase in the number of marriage events observed, and among remarriages there is a two-fold increase.

9 Note that first marriage and higher order marriage are defined on the basis of husband's marriage number

Potential Pairs Model Specification:

Similar to the characteristics used to define the dependent variable, there are four spousal characteristics that are included in mate-selection models – age, education, earnings, and marriage order. Those variables were selected because of their relevance to retirement and Social Security policy. Social Security rules indicate that age and earnings differentials directly affect the eligibility, benefit calculation, and claiming behavior of retired couples. Although education does not have direct implications for Social Security benefits or eligibility, it is a consistent predictor of mate selection (Mare, 1991; Pencavel, 1998; Qian, 1998). In addition, potential imperfections in the earnings measure make it important to include education in a mate-selection model. Including an education measure in the model allows low-earning, highly educated individuals to be distinguished from low-earning, poorly educated individuals. And finally, because the differences in spousal characteristics are likely to differ by marriage order, these models are estimated separately for first marriages and for higher order marriages.

This specification for modeling the likelihood of a match is much richer than the exponential distance function based on age and education used in DYNASIM. This specification, however, is more parsimonious than CORSIM and DYNACAN because it aligns on fewer dimensions. Characteristics such as race, fertility, and labor force participation at the time of marriage, which are included in CORSIM and DYNACAN, are not included in the models of union formation estimated in this paper because they do not

---

only.

have significant implications on Social Security eligibility, benefit calculation, or claiming behavior.

Age differentials are one of the most salient selection criteria in a mate-matching algorithm. Empirical evidence from the SIPP suggests that there is a nonlinear relationship between spousal age differences and the likelihood of a match. To accurately model this nuance, a combination of age splines and dummy variables are employed in the regression models. For both first marriages and remarriages, dummy variables are used to capture the most likely age differences for marriages. For first marriages, two dummy variables are used – one at an age difference equal to zero, and another at an age difference equal to one. Spline variables in the first marriage model separate the sample by the following age differentials: less than -7 years, -6 to -1 years, +2 to +7 years, and greater than +7 years. In the remarriage model, a single dummy variable is used to capture the baseline assumption of age difference – differences of zero, one, or two years. Spline variables separate the remarriage sample at the following age differentials: less than -8 years, -7 to -1 years, and greater than +2 years.

The microsimulation model used for this mate-matching exercise attempts to differentiate between highly educated and poorly educated individuals. Consequently, educational attainment is measured with a dummy variable that equals one if years of education are greater than or equal to 14, and zero otherwise. This limitation may reduce the descriptive power of the mate-matching process, and finer education detail is likely to be included in the future.

To account for potential differences in the joint characteristics of spouses by marriage order, potential pairs data files are created separately for men's first marriages and men's higher order marriages. In addition to estimating the models separately, each model has a dummy variable indicating whether the potential wife's marriage is her first or not. Thus, in the model for men's first marriages, a potential wife also on her first marriage is expected to be positively correlated with the likelihood of a match, whereas in the model for men's higher order marriages, a negative correlation would be expected.

Finally, a measure of earnings is also included in the model. Empirical evidence from the SIPP indicates that spouses have a tendency to have relatively similar earnings levels. To capture this economic homogamy, individuals are classified into sex-specific quintiles of average lifetime earnings (ALE), and the difference in quintiles is calculated.[10] The difference is specified as husband's ALE quintile minus wife's ALE quintile and ranges from –4 to +4. Both a linear and a squared term are included in the model to capture the quadratic relationship between ALE quintile difference and the likelihood of a match. A quadratic specification assumes that the direction of the difference in the ALE quintiles is not important. That is, marriages in which the man has higher earnings than the wife are just as likely or unlikely as marriages in which the man has lower earnings than the wife.

There are currently no cohort variables included in the model. There is some research, however, that the assortative mating patterns with regard to education have changed over time. This phenomenon is largely attributable to the increased educational attainment in

10 See the appendix for a description of how average lifetime earnings (ALE) are calculated.

the United States and the dependence on schools as marriage markets for nubile singles to find mates (Mare, 1991; Qian, 1998). Consequently, the joint distribution of spousal characteristics in future unions is assumed to remain the same as those observed in the mid-1990s.

Matching Algorithm:

The algorithm for mate matching starts with two pools of men and women to be matched in a given year.[11] The first step involves randomly sorting each of the lists of men and women. The next step cuts marriage candidates from their queue if the sizes of the pools are unequal. Excess "marriageables" have their marital status returned to their previous state and are sent back to the general population that will be at risk of marriage again in the following year. Removing excess individuals from the marriage queues creates two equally long, randomly sorted lists of men and women to be matched.

The mate-matching process presented here is male centric – each male finds a mate before proceeding to the next male. The low predictive power of the potential pairs model, however, suggests that a matching algorithm would take a very long time and may require several loops over the available females before a match is made. To mitigate this inefficiency, the matching process calculates a normalization factor that is then used to adjust the predicted probabilities.

11 Note that the mate-matching process described here applies only to nonimmigrants. Immigrants in the microsimulation model go through a separate mate-matching process based solely on difference in age.

For each man, a search across all remaining women calculates the normalization factor, which is set equal to the highest predicted probability of a match between him and all the potential women. A second pass through the list of women is when matches are actually determined. For each potential match, a random number is drawn and the predicted probability of the match is divided by the normalization factor. If the adjusted predicted probability is greater than the random number, the match is made. After the match is made, the female is removed from the list of females to be married that year and the algorithm proceeds to the next male to be married. The same steps are repeated until all males have been matched to all females.

The use of the normalization factor ensures that there will be a match made within the second cycle through the available women. The match that has the highest likelihood in the first cycle will have a likelihood of matching equal to one in the second cycle. This produces a similar effect as the methodology implemented in DYNASIM. In DYNASIM, a male searches over a random selection of 10 available women. If a match is not made on the first pass of those women, then the best match is assigned. In the mate-matching technique employed here, however, the number of women that men search over is *potentially* the entire set of remaining unmarried women. But because the location of the woman that would produce a "match with certainty" is randomly located in the queue, the actual number of women that each man searches over is also random. This technique creates a more randomized process than the one employed in DYNASIM, which arbitrarily limits the search to 10 women for each man before a match is made with certainty.

4. Model Results and Simulated Characteristics of Paired Couples

Regression results from men's first and higher order marriage models are presented separately in Table 1. Almost all of the coefficients are highly statistically significant. Many of the coefficients reveal expected correlations between husband and wife characteristics. Not surprisingly, these results suggest homogamous matchings. Individuals close in age, education, and historical earnings are more likely to marry than individuals who differ by those traits. Similar parameter estimates in columns one and two suggest that the relationship between individual traits and union formation does not differ significantly by marriage order.

Figures 1 through 6 present the results generated by those parameter estimates. In each figure a comparison to the SIPP is included. Figures 1 through 3 are for male's first marriages and Figures 4 through 6 are for male's second or higher order marriages. Figure 1 shows the distribution of simulated spousal age differences against the age differences of newlywed couples observed in the SIPP. Percentage of marriages is plotted on the y-axis and spousal age difference is plotted on the x-axis. The simulated values are based on the annual average distribution of the differences between 2002 and 2076. The distribution plotted for the SIPP is based on marriages that were formed between 1994 and 1996.

One is immediately struck by how well the simulated distribution matches the benchmark distribution in Figure 1. Because the SIPP has considerably fewer marriages than those produced over the 75-year projection period, there is slightly more variation in the distribution of the SIPP characteristics. The figure shows that both the benchmark

distribution and the simulated distribution have their peaks at +1. This peak indicates that approximately 15 percent of men's first marriages in the SIPP and over the simulated 75-year projection period are to women that are one year younger. The distribution is fairly even on either side of that peak and approximates a normal distribution.

Figure 2 shows the distribution of education differences produced by the stochastic mate-matching technique discussed above. This distribution indicates that there is strong homogamy by education, which is extensively supported in the research literature (Mare, 1991; Pencavel, 1998; Qian, 1998). The distribution of education differences produced in the simulation closely matches the distribution observed in the SIPP. In each, approximately three-quarters of the couples have the same education level, and regardless of which spouse has more education, the proportions with educational inequities are approximately equal.

Figure 3 shows the simulated distribution of differences in average lifetime earnings quintiles between husbands and wives. The distribution of each appears to be relatively symmetrically distributed, with more than a quarter of the couples marrying spouses that have the same relative economic ranking. Fewer than 5 percent of the marriages in the simulation and in the SIPP have extreme differences between the husbands' and the wives' economic status. This figure shows that the stochastic mate-matching approach produces slightly more marriages characterized by higher earning women (relative to their husbands) than the SIPP might suggest. This pattern may be the result of future compositional changes in marriage markets. As female earnings approach parity with male earnings,

there are likely to be ramifications on the characteristics of the matches produced. The overall congruency between the simulated differences in economic status and historically observed differences, however, is satisfactory. Matching along this dimension is particularly important because Social Security spousal benefits and workers' own benefits are affected by relative spousal earnings differentials.

Figures 4 through 6 present similar distributions as Figures 1 through 3, but they show the distribution of spousal differences among men's remarriages rather than first marriages. Figure 4 shows the simulated difference in ages between husbands and wives for men's higher order marriages and the similar differences seen in the SIPP. Those distributions differ significantly from the ones present in Figure 1 (for men's first marriages). The distribution from the SIPP for men's higher order marriages is much more varied than for men's first marriages and is not nearly as symmetrically distributed. The peak of the distribution occurs at +2 (indicating a marriage in which the man is two years older than the woman) and appears to be more heavily weighted toward the positive side of the distribution. The SIPP also shows a spike in the percentage of marriages in which the man is 20 or more years older than his spouse.

The simulated remarriage distribution, however, has much less variation and does not match the distribution observed in the SIPP as well as the distribution produced for first marriages. Overall, the simulated distribution suggests that older men remarry younger women. While this is generally the case, the SIPP also suggests that more simulated unions between older women and slightly younger men should exist. Another concern is

that the proportion of marriages produced where the husband is more than 20 years older than his wife is more than twice the size of the spike observed in the SIPP. Those concerns are somewhat mitigated by the fact that they are unlikely to have any significant effects because remarriages make up a small proportion of projected marriages.

Figures 5 and 6 are remarkably similar to Figures 2 and 3. Figure 6, the difference in average lifetime earnings quintiles among men's higher order marriages, shows a slightly more uneven distribution than the chart for men's first marriages. That pattern indicates that men in higher order marriages are more likely to marry women with lower earnings histories than men in their first marriage.

5. Conclusions

This paper provides an overview of methods available for matching new spouses in a dynamic microsimulation model and the particular methodologies used in other notable microsimulation models. Based on results from the CORSIM and DYNACAN models, it was determined that the stable marriage algorithm provides unsatisfactory results. This shortcoming forced CORSIM and DYNACAN to adopt a stochastic framework for matching spouses. The methodology employed by DYNASIM has always been based on a stochastic framework, although the specification of the exponential distance function is rudimentary. The mate-matching methodology presented here is also a stochastic-based technique that builds on the CORSIM and DYNACAN models and makes substantial improvements to the method used in DYNASIM.

While DYNASIM uses an exponential function based on differences by age and education, the technique shown here uses a potential pairs logistic regression model similar to the one employed in CORSIM and DYNACAN. The model presented here aligns spouse matches along age, education, and earnings differentials as well as by marriage order. The relatively parsimonious potential pairs model specification coupled with the use of a normalization factor to decrease the execution time for the matching algorithm produces an efficient matching technique for microsimulation models that closely replicates baseline distributions.

While there are significant similarities between the mate-matching technique described here and the one used in CORSIM and DYNACAN, the technique here appears to produce much better distributional properties relative to benchmark distributions. As Bouffard and others (2001) suggest, the inability to closely replicate joint-spousal characteristic distributions may result from the inclusion of poorly measured variables (such as fertility) in the mate-matching compatibility index estimation.

Over a 75-year projection period, the average joint-spousal characteristics of new marriages correlate extremely well with the baseline spousal characteristics. The ability to consistently replicate a fundamental relationship in microsimulation models is extremely important to producing reliable simulation results. Although the joint characteristics of spouses are an often-overlooked aspect of microsimulation models, the veracity of marriages produced has significant effects on the long-term projections of Social Security eligibility, benefit amounts, and collection duration.

References

Bouffard, Neal, Richard Easther, Tom Johnson, Richard J. Morrison, and Jan Vink. 2001. "Matchmaker, Matchmaker, Make Me a Match." *Brazilian Electronic Journal of Economics*. Vol. 4, No. 2.

Easther, Richard and Jan Vink. 2000. "A Stochastic Marriage Market for CORSIM." Strategic Forecasting Technical Paper: http://www.strategicforecasting.com/cgi-bin/filedesc.cgi?filename=REJVmarriage_101000.ps&path=REasther_paper/

Gale, D. and L.S. Shapely. 1962. "College Admissions and the Stability of Marriage." *American Mathematical Monthly*, 69:9-15.

Gusfield, Dan and Robert W. Irving. 1989. *The Stable Marriage Problem: Structure and Algorithms*. MIT Press: Cambridge, Mass.

Harris, Amy, Josh O'Harra, Ben Page, Kevin Perese, John Sabelhaus, Michael Simpson, and Joel Smith. 2002. "Overview of the Congressional Budget Office Long-Term (CBOLT) Policy Simulation Model." CBO Paper (forthcoming).

Johnson, Tom. 2000. "Stable Marriages vs. Optimal Marriages." Strategic Forecasting Technical Paper: http://www.strategicforecasting.com/cgi-bin/filedesc.cgi?filename=TJstable_102400.ps&path=TJohnson_paper/

Knuth, Donald E. 1976. *Marriages Stables*. Les Presses de l'Université de Montréal: Montréal, Quebec, Canada.

Mare, Robert D. 1991. "Five Decades of Educational Assortative Mating." *American Sociological Review*, 56:15-32.

O'Harra, Josh and John Sabelhaus. 2002. "Projecting Longitudinal Marriage Patterns for Long-Run Policy Analysis." CBO Technical Paper 2002-2.

Orcutt, Guy H., Steven Caldwell, and Richard Wertheimer II. 1976. *Policy Exploration Through Microanalytic Simulation*. Washington, DC: Urban Institute Press.

Pencavel, John. 1998. "Assortative Mating by Schooling and the Work Behavior of Wives and Husbands." *American Economic Review*, 88:326-329.

Qian, Zhenchao. 1998. "Changes in Assortative Mating: The Impact of Age and Education, 1970 - 1990." *Demography*, 35:279-292.

Zedlewski, Sheila R. 1990. "The Development of the Dynamic Simulation of Income Model (DYNASIM)." In Gordon H. Lewis and Richard C. Michel, Editors.

Microsimulation Techniques for Tax and Transfer Analysis. Washington, DC: Urban Institute Press.

Appendix A.

Construction of Average Lifetime Earnings

Average lifetime earnings are calculated on the basis of longitudinal Social Security earnings records linked to the 1996 SIPP panel. In the SIPP potential pairs logistic model, all the marriages start in 1994, 1995, or 1996. To ensure that the measure of average lifetime earnings captures only earnings prior to marriage, earnings from age 16 (whatever year that occurs in or 1951, whichever is greater) through 1993 are included. For all prior years, earnings are adjusted for inflation and real wage growth so that the calculated average lifetime earnings are in real 1993 dollars. This produces a measure that captures average lifetime earnings at one to three years before marriage. Finally, the measure of average lifetime earnings is calculated on the basis of Social Security earnings only. Any earnings above the Social Security taxable maximum and any earnings from noncovered employment are not included in the calculations.

Once calculated, the average lifetime earnings measure is used as a proxy for broad socioeconomic status. To do this, individual average lifetime earnings are ranked by sex and divided into quintiles. When categorizing individuals into quintiles, the sample that is included in the ranking is important to consider. The goal of the classification is to rank the men and women who are getting married in the years just before marriage relative to other nonmarried men and women. As such, the sample for the ranking includes SIPP observations with linked earnings records, for individuals who were never married or those who were married in 1996 but had their most recent marriage (or remarriage) in 1994, 1995, or 1996.

Table 1
Potential Pairs Logistic Model Results

	Men's First Marriages	Men's Higher Order Marriages
Intercept	-3.953 *	-3.647 *
	(0.070)	(0.080)
Wife's marriage number	0.973 *	-0.063
	(0.060)	(0.073)
Wife has higher education	-1.147 *	-1.054 *
	(0.043)	(0.120)
Wife has lower education	-1.104 *	-0.464 *
	(0.046)	(0.091)
ALE quintile difference[1]	-0.062 *	0.048
	(0.008)	(0.023)
ALE quintile difference squared[1]	-0.047 *	-0.084 *
	(0.004)	(0.010)
Age difference less than -6 spline	0.390 *	--
	(0.016)	
Age difference -6 to -1 spline	0.570 *	--
	(0.020)	
Age difference equal 0 dummy	-0.267 *	--
	(0.053)	
Age difference equal 1 dummy	-0.120	--
	(0.051)	
Age difference 2 to 7 spline	-0.216 *	--
	(0.010)	
Age difference greater than 7 spline	-0.238 *	--
	(0.007)	
Age difference less than -7 spline	--	0.250 *
		(0.018)
Age difference -7 to -1 spline	--	0.370 *
		(0.032)
Age difference equal 0, 1, or 2 dummy	--	-0.319 *
		(0.101)
Age difference greater than 2 spline	--	-0.107 *
		(0.007)
N	695,556	196,249
-2 Log Likelihood	57,674	10,917

Standard errors in parentheses, * = $p < 0.01$

1) Average Lifetime Earnings (ALE) quintile is calculated as husband's ALE quintile minus wife's ALE.

Figure 1
Distribution of Differences in Age Among Men's First Marriages:
Simulated Marriages 2002-2076 and SIPP Marriages 1994-1996

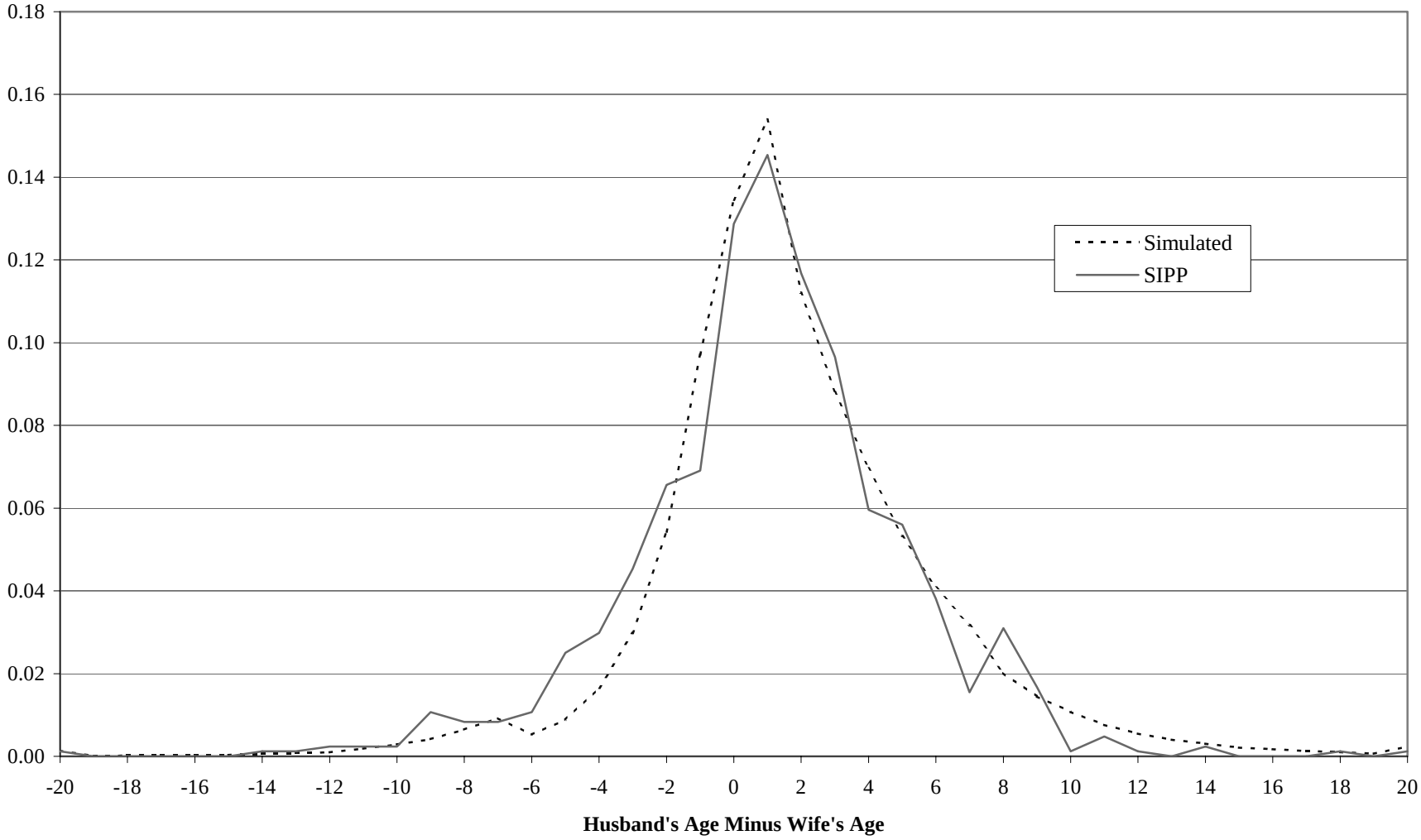

Figure 2
Distribution of Differences in Education Among Men's First Marriages: Simulated Marriages 2002-2076 and SIPP Marriages 1994-1996

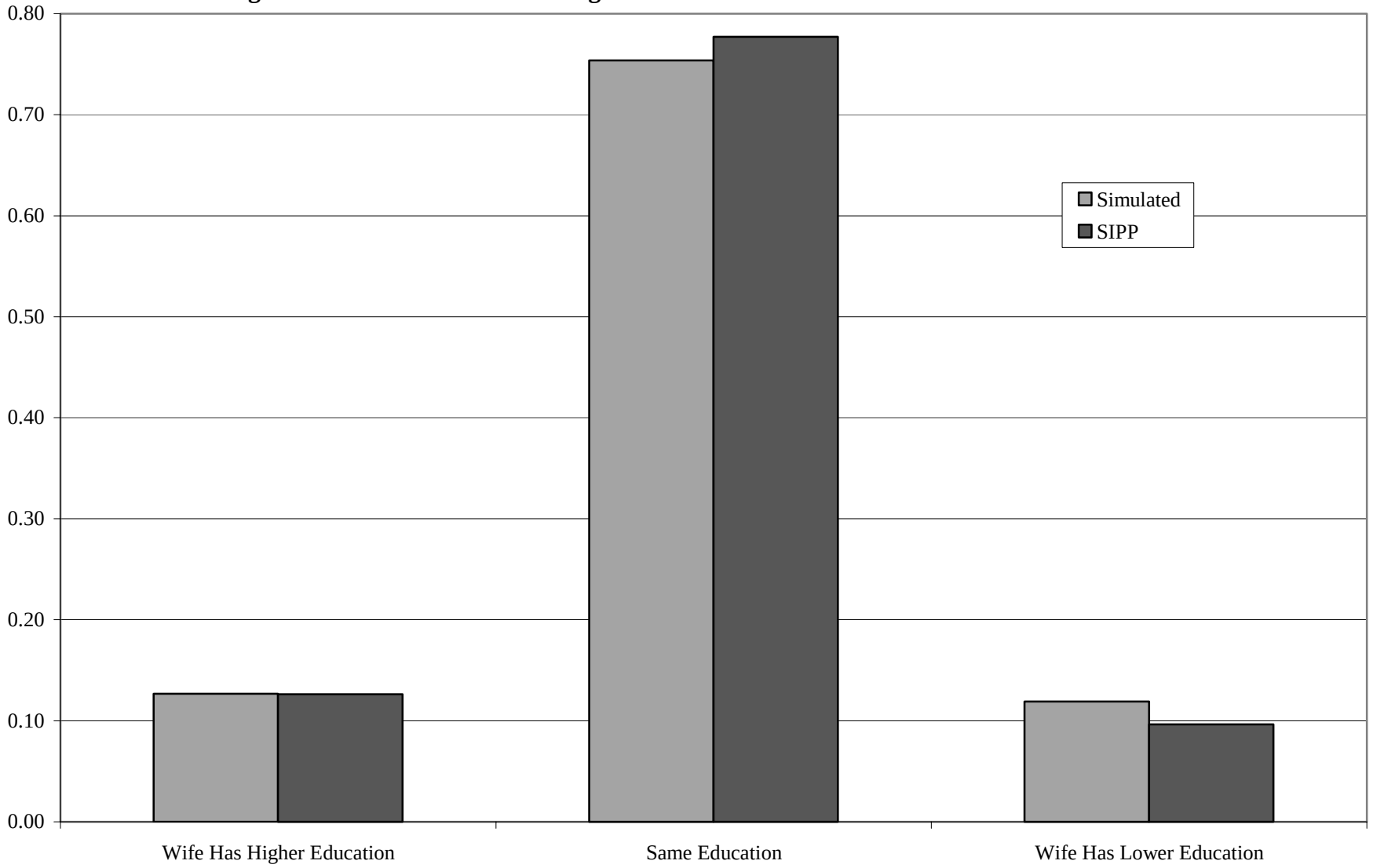

Figure 3
Distribution of Differences in Earnings Quintiles Among Men's First Marriages: Simulated Marriages 2002-2076 and SIPP Marriages 1994-1996

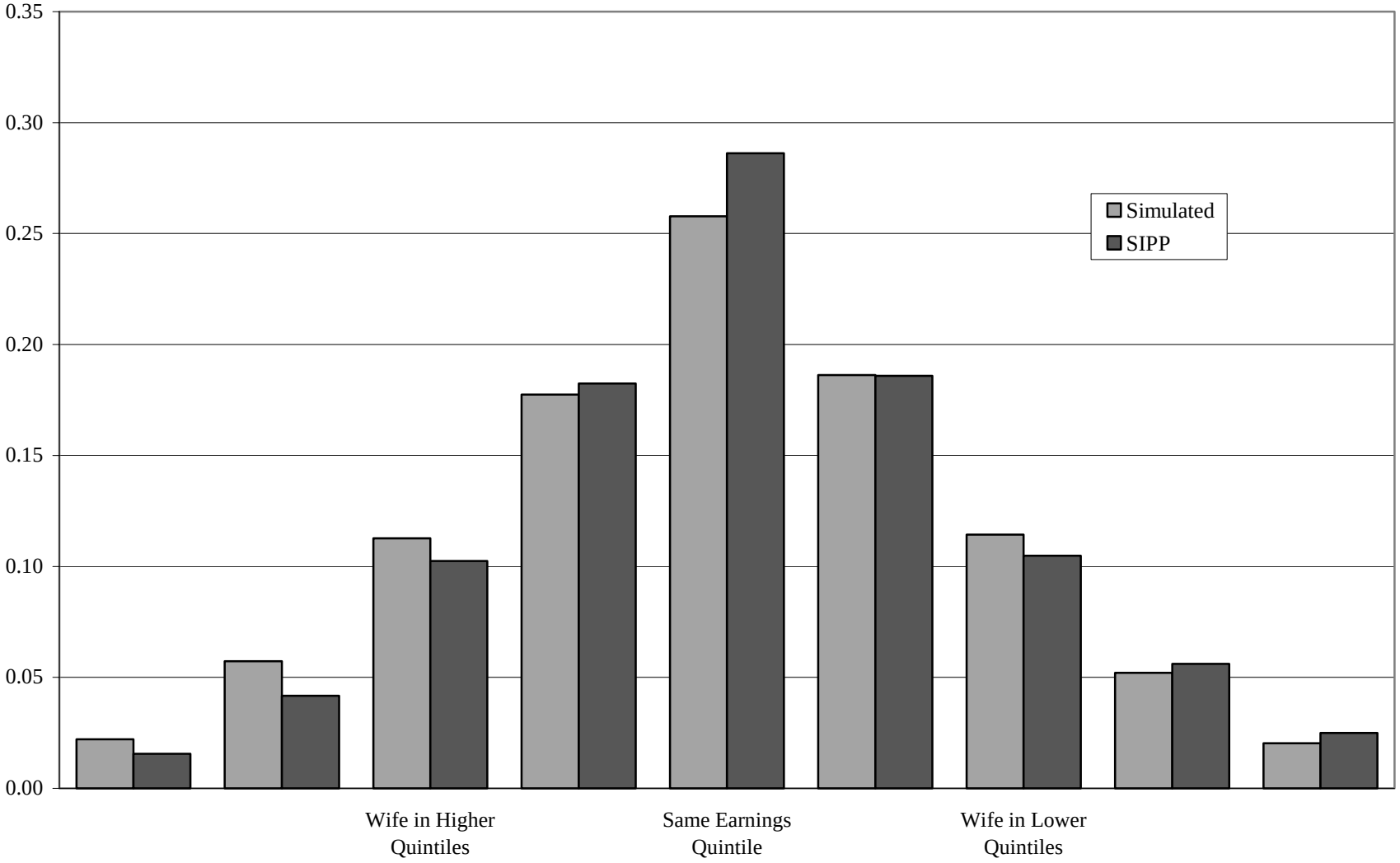

Figure 4
Distribution of Differences in Age Among Men's Higher Order Marriages: Simulated Marriages 2002-2076 and SIPP Marriages 1994-1996

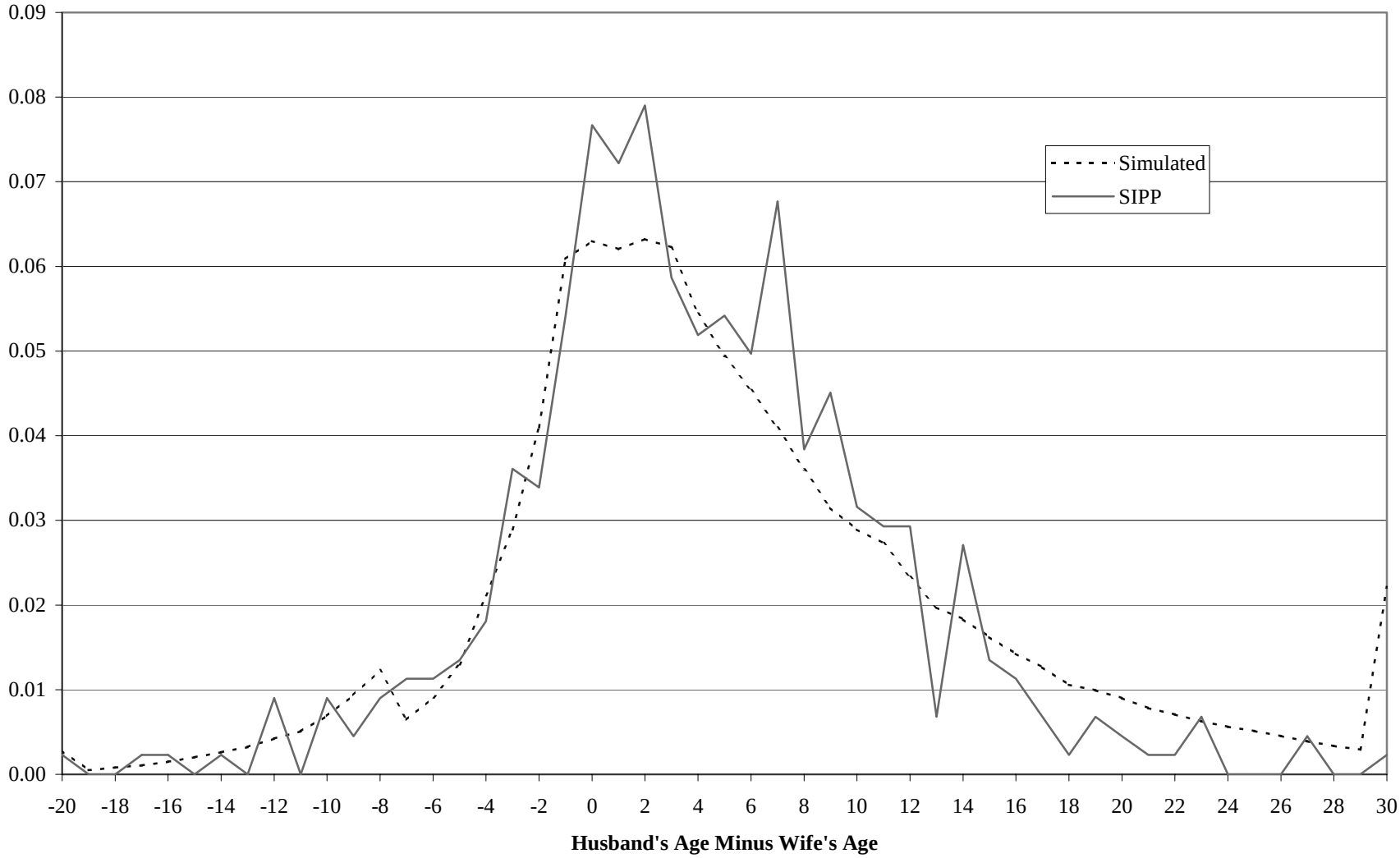

Figure 5
Distribution of Differences in Education Among Men's Higher Order Marriages: Simulated Marriages 2002-2076 and SIPP Marriages 1994-1996

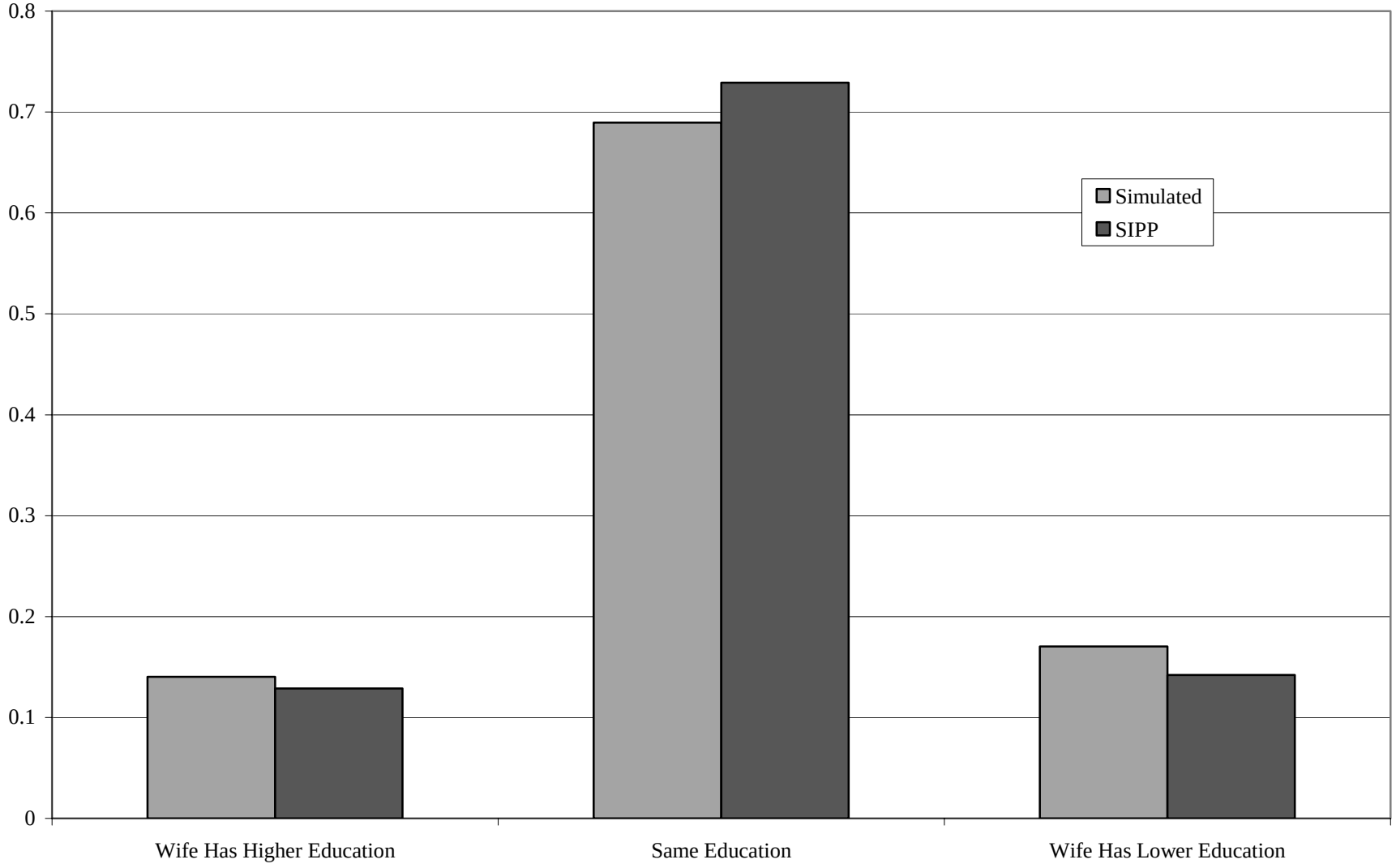

Figure 6
Distribution of Differences in Earnings Quintiles Among Men's Higher Order Marriages: Simulated Marriages 2002-2076 and SIPP Marriages 1994-1996