# The Marital Wage Premium in the Twenty-First Century:

*Do married men earn a higher wage rate, and if so, why?*

**Hans F. Kist, Yanqi (Halley) Hu**

*Dr. Marjorie B. McElroy*

---

*Honors Thesis submitted in partial fulfillment of the requirements for Graduation with Distinction in Economics in Trinity College of Duke University*

Duke University
Durham, North Carolina
2010

## Table of Contents

**Acknowledgements**

Special thanks to Dr. Marjorie McElroy and our fellow classmates in Econ208S and Econ195.14 for suggestions and guidance throughout the past year of research. We also wish to express gratitude to the interns in the Perkins 226 computer lab for assisting us with conducting empirical work in STATA. This paper would not have been possible without these generous people and their valuable advice.

**Abstract**

Married men have historically earned higher wages than single men. One of the most prominent explanations for this phenomenon is the theory of intra-household specialization. However, the marriage premium was found to be decreasing up until the early 90's. In our paper, we have re-examined the wage premium using data from the National Longitudinal Survey of Youth 1979. Our analysis focuses on an early period (1990-1992) and a later period (2002-2006). Our results suggest that the marital wage premium has actually been increasing over time, but that specialization does not do an adequate job of explaining this result.

## Section I – What is the Marital Wage Premium?

One of the more persistent findings in labor and family economics is that married men tend to earn more than single men, even when age, prior work experience, and many other personal characteristics are controlled for. This difference in earnings has generally been quite large. For example, studies in the 1970's and 1980's suggested that married men tended to earn between 10% and 40% more than their single counterparts, and that marriage was as important as race, firm size, and union membership in determining a man's wage (Korenman and Neumark, 1991). However, evidence suggests this difference in earnings has also fallen over time. Some estimates suggest that this difference fell almost 20% from 1968 to 1988 (Blackburn and Korenman, 1993). Other studies find that the difference in wages fell from 11% in the late 70's to 6% in the early 90's, suggesting an even more drastic decrease in earning differences of almost 50%(Gray, 1997). There are two very interesting and obvious questions to ask. One, why does this difference in earnings between married and non-married men, recognized by economists as 'the marital wage premium', exist? And two, why has it seemingly been on the decline over the last generation?

Our paper will try to answer these questions using data from the 1979 National Longitudinal Survey of Youth (NLSY79). Focusing on the survey years 1990-1992 and 2002-2006 will let us determine whether or not the 'marital wage premium' has continued to decline in the last ten to fifteen years, and will help us try to uncover the reasons behind this change.

Section II summarizes past literature on the marital wage premium and discusses how this study contributes to the body of research in this field. Section III introduces the economic theories behind the wage premium, as well as our predictions for the current premium based on

each theory. Section IV explains the NLSY79 data and how we plan on utilizing it in more detail. Section V gives the econometric specifications we plan to use, and will also discuss how we can to use our specifications to test the theories explained in section III. Section VI is devoted to presenting the results from our regressions, and Section VII is devoted to analyzing and explaining these results. The limits of our results and any possible endogeneity problems are discussed in Section VIII. We will conclude our paper in Section IX with a summary of our main findings and a discussion of future research ideas that can add to the 'marital wage premium' literature.

## Section II – The Marital Wage Premium and the Theories of Specialization and Selection: a Review of Previous Literature

Previous studies on the marital wage premium have primarily been focused on the size and the cause of the wage differential. A prominent explanation used by economists is the theory of specialization. Pioneered by Gary Becker, this theory suggests that specialization within a household can lead to increased welfare, and for the household member specializing in market activity, it can lead to higher wages. Since the conception of this theory, much empirical work has been done which suggests that household specialization is indeed an important factor in determining the size of the wage premium. For example, in a 1991 study by Korenman and Neumark, it was found that married men's wages increased with each additional year married, which is highly consistent with the explanation that the marital wage premium is due to specialization within households.

However, not all studies find strong evidence for household specialization being the prime determinant of the marital wage premium. Some studies find that the selection of males

into marriage based on personal characteristics that are positively correlated with higher wages is as important if not more so than specialization. Many papers written since Korenman and Neumark's 1991 study have attempted to more accurately test these competing theories of specialization and selection.

We have summarized a number of such studies in the table provided. In addition, the table also includes two more studies that have looked at the change in the marital wage premium over time. If specialization is a factor in the marriage premium, then we would expect the premium to change over time as social conditions that effect household specialization change. The size of this change is another interesting and important empirical question.

| Author(s) | Title | Year of Publication | Data source | Years observed | Cross sectional/Longitudinal | Conclusions | |
|---|---|---|---|---|---|---|---|
| **Ginther and Zavodny** | *Is the Male Marriage Premium Due to Selection?* | 2000 | National Longitudinal Survey of Young Men , 1980 Census 5% Public Use Microdata Sample | 1970, 1976 | cross sectional | 90% marriage premium remains after control for selection | evidence for the existence of the marital wage premium |
| **Hersch and Stratton** | *Household Specialization and the Marital Wage Premium* | 2000 | National Survey of Families and Households | 1987-88, 1992-94 | longitudinal | neither selection nor specialization causes marital wage premium | causes behind the marital wage premium |
| **Neumark and Korenman** | *Does Marriage Really Make Men More Productive?* | 1991 | National Longitudinal Survey of Young men | 1976-1980 | both | attributed marriage premium to specialization. Specifically, the premium accumulates through years married. | |
| **Chun and Lee** | *Why do Married Men Earn More* | 2001 | Current Population Survey March Supplement | 1999 | cross sectional | specialization is the driving force behind the martial wage premium | |
| **Blackburn and Korenman** | *The Declining Marital-Status Earnings Differential* | 1993 | Current Population Survey, Census of Population | 1970, 1980 | cross sectional | Changing roles within marriage, and changes in human capital accumulation within marriage accounted for much of the decrease | causes behind the decline in the marital wage premium |
| **Gray** | *The Fall in Men's Return to Marriage* | 1997 | National Longitudinal Survey of Young Men, National Longitudinal Survey of Youth | 1976-1980, 1989 to 1993 | both | productivity enhancing attributes of marriage have substantially decreased, leaving the remaining wage premium to be described by the selection of high wage men into marriage | |

For our study, we are re-examining this wage differential using the most recent data collected from the NLSY79 (National Longitudinal Survey of Youth 1979). We will restrict our analysis of this data to two time periods: one from 1990 to 1992, and the other from 2002 to 2006. This combination of early and later data will allow us to contribute to the research on how the wage premium has changed over time.

In addition, two important differences in our methodology will let us better test for the effects of specialization. One is using information on the wage rate of a person's spouse. It may be the case that the adult male is not always the household member specializing in market activities, and so earlier studies that assume this to be the case may underestimate the effect of specialization on the marital wage premium. In addition, we will control for the probability of divorce in our analysis. Anticipated divorce or separation will lead to less specialization. Thus studies that don't control for this would likely find it hard to accurately estimate the importance of specialization, as they would have difficulty determining how much specialization is actually occurring.

These two differences in our methodology will allow us to make original contributions to the research in this field. In addition, by using the most recent data from the NLSY79 we will be getting more current estimates of the marital wage premium, which will benefit researchers in this field as well.

We are now ready to move into a more detailed explanation and discussion of the theories of specialization and selection.

## Section III – Specialization, Selection, and Favoritism: Three Competing Explanations for the Marital Wage Premium

As already mentioned in our literature review, there are two main theories (and sometimes a third) most often used to explain the marital wage premium. This section will explain these theories in greater detail, and then discuss what predictions can be made from these theories regarding our research question. That is, what do these theories say about the state of the marital wage premium today. Do they suggest the premium has grown, decreased, or stayed roughly constant?

### *i) The Theory of Specialization*

The first theory to be discussed is the theory of specialization. It is perhaps the theory most often used to explain the marital wage premium. This theory is attributed to Gary Becker (1981), and suggests that the marital wage premium is due to differences in productivity between married and non-married men. The theory can be explained as follows.

Becker holds that many products people actually gain utility from are a combination of items bought in the market and time spent in the home modifying these items into a final product which gives utility to its owner(s). For example, most people only gain utility from the food they buy at the supermarket once it has been cooked and combined together to make an actual meal. Perhaps beautiful furniture bought from a store is most enjoyed once time has been taken every day at home to remove all the dust and dog hair. The main idea is that in order to enjoy many products, time must be spent both in market activities and home production. Someone has to make the money to buy the food from the grocery store, and someone has to take the time to actually cook the meal.

Becker suggests that our efficiency in doing these activities depends on the amount and type of "human capital" we have. For simplification purposes Becker divides human capital into two categories: market related capital and household related capital. This capital can be thought of as the skills, experience, and knowledge we need to help us perform tasks in the market and at home. Clearly, the type of human capital needed to excel in the market is not always the same type of human capital needed to excel at in-home production. A very successful manager or investor has likely spent a lot of time acquiring market-related human capital. They may have spent several years in business school or getting an advanced degree. They have likely learned intricacies of their job that only years of experience could have taught them. However all the skill and knowledge they possess about managing people or investing money would not help them cook better tasting meals or be a better housekeeper. Likewise, someone who has spent years cleaning the house and cooking dinner while their spouse is at work probably can make much better tasting food and can perform other household duties faster and more effectively than another person without similar experience. Clearly however, the ability to cook great tasting meals does not translate into the ability to go into the market and find a high paying job.

The basic idea, therefore, is that people expend resources (time, money, etc) to gain human capital, but they have to choose which type of human capital they want to procure (market related capital or household related capital). A single person will likely invest in both market and home related human capital, as he will be responsible for both buying items in the market and spending time at home using those items to produce the final good he or she cares about. However a married couple can "specialize" their investments so that one of them becomes more effective at in-home production and the other becomes more effective at market related activities. They can then allocate the time spent at home and in the market in an efficient manner,

and jointly they should be able to produce more utility for each person than they could if they both were single. For more details on this theory, and specifically on why this ability to increase utility always holds (given certain assumptions), one should read chapter two of Gary Becker's book "A Treatise on the Family" (1991).

Becker's specialization theorem implies that the spouse who specializes in market related human capital and activity will tend to earn a higher wage rate than a comparable single person who cannot procure a similar amount of market related human capital, or who cannot spend an equal amount of time working in the market. This is because the additional market related human capital will tend to result in a more productive worker. Because men have historically been the ones to work in the market, and women have typically stayed at home, we would expect that married men would invest more in market related human capital. Similarly, married women would invest more in home related human capital. It then follows from Becker's specialization theory that married men should on average earn more than single men when other factors such as age, experience, and natural ability are controlled for.

## *Predictions for Today:*

If this theory of specialization is in fact the driving force behind the difference in earnings between married and single men, what changes should we expect to be seeing in the marital wage premium?

Many cultural changes would suggest that the wage premium should be decreasing. For example, more and more housekeeping activities are being overtaken by the market. A single man doesn't have to clean his house or apartment, but can hire a maid. While maids are certainly not a development of the modern age, increasing levels of per capita income suggest that more

single men are able to outsource such household activities to the market. While not all single men are likely to be preoccupied by keeping their living space clean, rising levels of personal income would mean that they can outsource household activities that they do care about, such as cooking. The obvious example here is that more people could eat out at restaurants instead of expending time and effort making their own meal. In addition to the effect that rising levels of income and prosperity have on the ability to outsource household activities to the market, technological inventions that make every day activities easier can also have an important effect.

Secondly, more and more women are spending significant amounts of time in the workforce, both for social and financial reasons, suggesting that people are not specializing to the same degree as before. Thus we might expect that the gains to specialization are not be what the once were.

A somewhat related change in society is that more and more women are the higher earning member of a marriage or relationship. If more women in a marriage have a comparative advantage in market activities, we would expect that more and more men are actually spending time specializing in household related capital rather than in market related human capital. We would expect this to have a negative effect on the male marital wage premium, as males may not be all specializing in the market now. However, this last point does not mean that specialization is less important or effective in raising wages than before. People may still be specializing; more of them just may be specializing in something opposite of what we would traditionally expect. While this trend would suggest that the male marital wage premium is probably smaller than before, it makes it harder to make conclusions about the importance of specialization in determining the wage premium.

These factors suggest that the marital wage premium should have continued to decrease in the past ten to fifteen years since the study by Gray (1997). However, divorce rates have also fallen over that span (National Center for Health Statistics, 2009). If the probability of divorce is high, then couples may not have as strong an incentive to specialize, since they might have to resume being single, and would have to start working in both the market and at home again. Thus we would expect a high probability of divorce to lead to fewer gains from specialization. Conversely, if divorce rates have significantly fallen on an aggregate level, then individual probabilities of divorce have likely also been decreasing, and other things being the same, we would expect the gains from specialization to increase.

The relative importance of divorce probabilities and societal changes causing a decrease in market gains from specialization is an empirical question that we hope to shed some light on in later sections. We can however still make some predictions about the marital wage premium and how it has changed based on what the theory of specialization says. We have done this in Table 2, which summarizes our overall predictions for the effects that the following variables would have on the marriage premium (when they are all included in a regression together with a reference group of never married men): *a currently married dummy variable, a cumulative years married variables, a cumulative years divorced variable*, a *probability of divorce term multiplied by our years married variable, and finally a spousal wage variable multiplied by our years married variable.*

The table also predicts how these effects have likely changed over time. We look at the interaction of spousal wages and divorce probabilities with cumulative years married because we expect the effect of specialization on wages to show itself in a cumulative years married term, as this term will signal the amount of time an individual could have been gaining from extra

investment in market related human capital. Thus these interaction terms will tell us how these variables affect the degree of specialization occurring in a household.

Specifically, if we were to regress an individual's wage on the following variables, we would expect the signs and magnitudes of the coefficients to be as presented in the table (again assuming our reference group is never married men). This of course assumes that specialization is indeed the driving force behind the marital wage premium.

Table 2 – Predictions from Specialization

| **Explanatory Variables:** | **Sign of Coefficient at time of previous studies** | **Predicted Sign of Coefficient Now** | **Predicted Change in Magnitude** |
|---|---|---|---|
| Currently Married Dummy | Close to Zero | + | Unclear |
| Years Married | + | + | Decreased |
| Years Separated/Divorced | - | - | Decreased |
| (Probability of Divorce) x (Years Married) | - | - | Unclear |
| (Spouse's Wage Rate) x (Years Married) | - | - | Unclear |

Most of these predictions come directly from effects we have talked about in the preceding paragraphs. Specialization says that the wage premium arises out of differences in investments in human capital. Thus we would expect the benefit of marriage to be something that accrues through time as one continues to accumulate more and more market related capital. It follows that the coefficient on years married should be positive, and should dominate the effect of current marital status. The predicted negative coefficient on years separated/divorced can be justified as follows. Someone who has been married has built up a surplus of market related human capital when compared to a man who has never married and never been able to specialize.

When that person gets divorced they have to change their human capital investment patterns and stop specializing in market related activities. From then on, they will start losing some of their surplus market capital to depreciation. This explains the negative prediction for the coefficient on years separated/divorced. We have already talked about how higher wages for one's spouse and higher probabilities of divorce are likely to affect investment patterns in human capital and thus the wage premium. Also, when the probability of divorce is controlled for, we would expect investment patterns of single and married men to be more similar than in the past due to several societal trends already mentioned. This explains the predictions in the right-most column.

Later we will be able to use these predictions to help see what explanation or theory of the marital wage premium is supported most by our data. We are now ready to move on to a discussion of the selection hypothesis, which was already mentioned in section II of this paper, and which after the theory of specialization is the most popular explanation of the wage premium. We will also mention a third, less popular explanation of the wage premium.

### *ii) The Selection Hypothesis (and the theory of favoritism)*

The other primary theory used to explain the marital wage premium is the idea that there are personal characteristics that may be unobserved, but that are valued in both the marriage and job markets. This theory suggests that people are selected into marriage based on these personal characteristics, which are of course correlated with higher wages. We expect therefore that these personal characteristics will be represented to a higher degree in the group of married men than in the group of single men, and since these characteristics are positively correlated with higher wages we would also expect to find that other things being equal, married men as a group tend to earn more than non-married men.

While the selection hypothesis and the theory of specialization tend to be the most popular theories in explaining the marital wage premium, there is also a third theory worth mentioning. This theory suggests that married men earn more than non-married men due to favoritism. Specifically, employers unfairly discriminate against non-married men, so that even though married and non-married men may be equally productive, the married man is more often given the promotions and raises.

### *Predictions for Today:*

Predictions from these theories are much less straightforward than predictions based on the theory of specialization. It is possible that as the types of jobs available to people change, so do the personal characteristics that make someone desirable in the job market. Perhaps then the characteristics that make someone an attractive marriage partner are no longer associated with higher wages. We can't be sure about what these characteristics are, especially if some or many of them are unobserved to begin with. On the other hand, it could just as easily be the case that selection into marriage and high wages are even more closely related than before. For example, we might expect that as the average age of marriage increases, people are spending more time searching for a mate, and so they are more likely than before to choose someone who has the relevant set of unobserved personal characteristics.

While it is therefore somewhat difficult to make predictions beforehand about the effect of the selection hypothesis on the marital wage premium, if the relevant personal characteristics are all fixed over time, we should still be able to control for them by using a longitudinal econometric specification (see Section V). Thus, we can still test the theory of selection in a general manner, but we won't be able to use the data to explain in detail how the selection

process may have changed over time (i.e. what personal characteristics are selected for now vs. in previous decades).

Nevertheless, what predictions we can make from the selection hypothesis are given in the upcoming table below. We would still think that marital status is correlated with higher wages. As the selection hypothesis gives us little reason to think that the bulk of the premium should accrue through time, we would also think that the number of years an individual has been married is not very important.

Again, the table predicts signs of coefficients for the same variables used in table 2, assuming again that our reference group is never married men. As we don't have any way of knowing how the selection process has changed over time, we can make no clear predictions on how the magnitude of our coefficients should change over time.

Table 3 – Predictions from the Selection Hypothesis

| **Explanatory Variables:** | **Sign of Coefficient at time of previous studies** | **Predicted Sign of Coefficient Now** | **Predicted Change in Magnitude** |
|---|---|---|---|
| Currently Married Dummy | Positive | Positive | Unclear |
| Years Married | Should Have Little to no Effect on the Wage Premium | Should Have Little to no Effect on the Wage Premium | Unclear |
| Years Separated/Divorced | Should Have Little to no Effect on the Wage Premium | Should Have Little to no Effect on the Wage Premium | Unclear |
| (Probability of Divorce) x (Years Married) | Should Have Little to no Effect on the Wage Premium | Should Have Little to no Effect on the Wage Premium | Unclear |
| (Spouse's Wage Rate) x (Years Married) | Should Have Little to no Effect on the Wage Premium | Should Have Little to no Effect on the Wage Premium | Unclear |

It should be clear from the above table, that the theory of specialization provides a much greater and more precise set of predictions than does the selection hypothesis. Nevertheless, these predictions will be useful later on when we try to find empirical evidence for the different economic theories behind the wage premium.

On the other hand, if favoritism is the driving force behind the marital wage premium, then our prediction of how the marital wage premium has changed over time depends on whether or not we think society is more or less prejudiced now than in previous decades. The answer to this is not immediately clear. While we might think that in general people are less prejudiced, it is much harder to verify this with data. Furthermore, we are looking at a very specific type of prejudice that is particularly hard to measure through currently available data.

In conclusion, the theory of specialization suggests that several societal trends have caused the marital wage premium to decrease over time, though declining divorce rates may act in the opposite direction, causing upward pressure on the premium. While the process of selection may have changed due to changing job opportunities or changes in the way spouses are chosen, it's not clear if it would have caused the marital wage premium to increase or to decrease. We have a set of predictions from the theory of selection, but while these predictions may be useful, they are nonetheless very limited. However, by differencing away personal fixed effects in our econometric model, we can hopefully determine whether or not the selection hypothesis has become more or less important over time.

It is not clear what predictions we can draw from the theory of favoritism. Also, while some have attempted to test the theory of favoritism (Neumark, 1991), doing so is typically much more difficult than testing either of the alternative theories. Thus our paper will primarily

focus on the theories of specialization and selection. We will now move on to a discussion of our data, before attempting to determine the size and causes of the marital wage premium.

## Section IV – Data Selection from the NLSY79

As stated before, the data we will be using come from the 1979 National Longitudinal Survey of Youth. This is a nationally representative panel data set that follows 12,686 individuals born between 1957 and 1964. From 1979 to 1994, individuals in the data set were interviewed every year, and have been interviewed every other year from 1994 onwards. We will use data from the years 1990-1992 and 2002–2006. Within each of these time spans, there were three surveys given: in 1990, 1991, and 1992, and in 2002, 2004, and 2006. There are several reasons for picking these sample years. The earlier period will give us a benchmark from which to make conclusions about how the marital wage premium has changed over time. The later period utilizes the most recent data from the NLSY79, and will allow us to get the most up to date estimate of the marital wage premium.

However, there are some drawbacks to using these two periods from the same survey. For one, the people in our sample will be 10 to 15 years older in our later sample than in our early sample. This means that we won't be able to perfectly see how the wage premium has changed over time for individuals of roughly the same age, though we can control for age to some extent in our econometric specification. Ideally we would be able to compare people of similar ages across time, however we have to work with the data that's available.

While having an older sample may be bad for comparison purposes, for simply estimating the current marital wage premium, there are benefits to having an older group of

people. In early adulthood, employee turnover is relatively high, as many people go through a period of job matching. In this period, people change jobs relatively often in order to gain experience and ultimately find the job they're best suited for. As theory holds that people will ultimately be matched with employers that value their services the most, their wages in this final job are arguably more representative of their true productivity and potential as workers.

The period of job matching varies, but typically someone in their forties is very likely to have settled into a long term job. Because the ages in our sample range from 37 to 49 over the period 2002-2006, we hope that the wages we will be looking at are better measures of lifetime productivity than if we had used a younger sample.

Moving on to the actual method of selecting the data we used for our regression analysis, the sample we will use in both periods was selected from an overall population of 12,686 people, as already stated. In each period, we threw out any observations from the original 12,686 that were not male. In addition, we dropped any Black and Hispanic observations from our sample. This decision was made because the convention in much of the previous literature has been to focus on white males. When we selected for sex and race we were left with 3790 observations in both the 1990-1992 period and the 2002-2006 period.

In addition, we wished to look at the following variables to conduct our analysis (presented on the next page). The reason we chose these variables will be discussed in section V.

After limiting our sample to non-Black and non-Hispanic males, we also threw out any observations that did not have information recorded on all of the variables in the table below for all sample years for the period we were interested in. For example, if someone lacked information for one of the variables in the table in 1991, that person would not be included in our

regressions focusing on the period 1990–1992. It is still possible that he was included in the regressions for the later years, as long as he had all the information we needed in those years.

| **Variable We Wish to Explain** | **Explanatory Variables** | **Control Variables** | **Variables Used to Estimate Probability of Divorce** |
|---|---|---|---|
| Hourly Wage | Marital Status | Age | Who Interviewee lived with at age 14 |
| | Yrs Married, $YrsMarried^2$ | Education(highest grade comp) | # of spouses/partners interviewee has lived with |
| | Yrs Separated/Divorced, $\ldots^2$ | Religion Raised In | First marriage before first birth dummy |
| | Duration of First and Second Marriages | Weeks Worked per Year | # Children considered ideal for a family |
| | Duration of First and Second Periods of Divorce | Hours Worked per Week | # Children actually in the family |
| | | Number of Children | Spouse's Income & Wages |
| | | Region of Residence | Sex of First Three Children |
| | | Urban or Rural Residence | Years Married |
| | | Union Membership | Education(highest grade completed) |
| | | Occupational Industry | Religion |
| | | Years of Work Experience | Age |
| | | Spouse's Wage | |
| | | Probability of Divorce Over Next Two Years | |
| | | | |

Ultimately, after throwing out all the observations that lacked records for the variables in the above table (as well as discarding observations with wages in the top and bottom 2% of our sample), we were left with 813 observations for the 1990-1992 period, and 634 observations for the 2002-2006 period. We decided to drop these wages to get rid of outliers, or individuals with wages that seemed highly out of place. Now we should explain some variables that needed to be

created from variables in the data set we originally downloaded from the NLS (National Longitudinal Surveys) website.

One of the variables we needed to create was a cumulative years married term. In each round of the survey we had information on the start and stop dates of an individual's first two marriages. We also had the start date for a third marriage if there was one. We dropped any observations who had been in a third marriage, as we had no way of knowing if they were divorced in that marriage or not. These variables allowed us to create variables specifying the duration on an individual's first two marriages as well as their first two periods of divorce. We could add these variables to get the cumulative number of years an individual had been married or divorced. Together, these variables told us a person's marital history.

The cumulative work experience variable was generated from information about how many weeks each person had worked since the last interview. Specifically, we added up the number of weeks worked since the last interview for each interview from 1979 to the year of interest to generate our cumulative work history variable. If this information (number of weeks worked since last interview) was not available for an individual in any survey year from 1979 up to the last date of the period being analyzed, they would have had an incomplete cumulative work experience variable, so in the tables above, we threw out any people who lacked such information.

We also ended up generating an estimate for spousal wages from information on the spouse's income from wages and salary in the past year, and the number of hours they worked during the past year.

The way we construct our estimated probability of divorce variable is more complicated and involves regressions. Thus the creation of this variable will be dealt with in our next section, where we discuss our econometric specifications.

Finally, tables of some summary statistics for our selected sample of individuals are given below. There is one table for each sample period (90-92 & 02-06). These numbers are averages across all individuals and survey rounds within the given sample period.

Following the presentation of our tables of summary statistics, we will move on to section V of this paper, where we explain how we will use our data to test if and why the marital premium has changed over the previous fifteen to twenty years.

## Summary Statistics from 1990-1992

| Variable | Mean | Std. Dev. |
|---|---|---|
| *Currently Married* | 0.5150 | 0.4999 |
| *Currently Divorced/Separated* | 0.1021 | 0.3028 |
| *Cumulative Years Married* | 3.7176 | 4.0852 |
| *Cumulative Years Divorced* | 0.4278 | 1.4693 |
| *Hourly Wage Rate (cents)* | 1142.3150 | 451.1132 |
| *Natural Log of Hourly Wage Rate* | 6.9607 | 0.4131 |
| *Age* | 29.6417 | 2.3574 |
| *Children* | 0.6769 | 0.9580 |
| *Education(highest grade completed)* | 13.5080 | 2.3845 |
| *Hours Worked in Past Calendar Year* | 2250.9910 | 590.5274 |
| *Weeks Worked in Past Calendar Year* | 49.1997 | 7.1624 |
| *Raised in a Religious Household* | 0.9533 | 0.2111 |
| *Estimated Probability of Divorce in the Next Two Years* | 0.0522 | 0.0444 |
| *Actually Get Divorced in the Next Two Years* | 0.0344 | 0.1824 |
| *Wage of Spouse (dollars)* | 5.2785 | 8.5691 |
| *Currently Have a Higher Earning Spouse* | 0.1636 | 0.3700 |

## Summary Statistics from 2002-2006

| Variable | Mean | Std. Dev. |
|---|---|---|
| *Currently Married* | 0.7298 | 0.4442 |
| *Currently Divorced/Separated* | 0.1356 | 0.3425 |
| *Cumulative Years Married* | 13.6483 | 7.9889 |
| *Cumulative Years Divorced* | 1.5201 | 3.6440 |
| *Hourly Wage Rate (cents)* | 2367.9410 | 1285.7480 |
| *Natural Log of Hourly Wage Rate* | 7.6157 | 0.6206 |
| *Age* | 42.7518 | 2.7720 |
| *Children* | 1.6272 | 1.3057 |
| *Education(highest grade completed)* | 13.8586 | 2.4829 |
| *Hours Worked in Past Calendar Year* | 2379.3430 | 581.1705 |
| *Weeks Worked in Past Calendar Year* | 50.8160 | 4.8633 |
| *Raised in a Religious Household* | 0.9558 | 0.2055 |
| *Estimated Probability of Divorce in the Next Two Years* | 0.0242 | 0.0192 |
| *Wage of Spouse (dollars)* | 10.7706 | 15.6613 |
| *Currently Have a Higher Earning Spouse* | 0.1551 | 0.3621 |

## Section V – Econometric Specifications: Methods for Testing for the Marital Wage Premium and its Causes

In order to estimate the marital wage premium we need to regress some measure of hourly wage on marital status, along with other control variables. We will run many different types of regressions, but they will all be variations on three basic specifications. First, we suppose the natural log of an individual's hourly wage is given by the following:

1) $$\ln(W_{it}) = \beta_0 + \boldsymbol{\beta_1} \cdot \boldsymbol{Marital\ Status_{it}} + \boldsymbol{\beta_2} \cdot \boldsymbol{X_{it}} + \boldsymbol{A_i} + \varepsilon_{it}$$

Where $W_{it}$ is the person's hourly wage rate, $\beta_0$ is a constant, $\boldsymbol{\beta_1}$ is a vector of coefficients, and $\boldsymbol{Marital\ Status}$ is a vector with several different marital states (ex: Married, Separated, Divorced, etc). $\boldsymbol{X_{it}}$ is a vector of observed personal characteristics (age, education, marital history, work experience, occupation, etc), and $\boldsymbol{A_i}$ is a vector of unobserved personal characteristics that is constant over time. The error term is denoted by $\varepsilon_{it}$. In all our regressions we will allow for the error term to be heteroskedastic.

Note that because $\boldsymbol{A_i}$ in the above specification is unobserved we have to first difference $\boldsymbol{A_i}$ away before we can control for these unobserved characteristics. Thus it will be beneficial to look that the second basic specification we will use (the longitudinal specification):

2) $$\ln(W_{it}/W_{i\cdot}) = \boldsymbol{\beta_1} \cdot (\boldsymbol{Marital\ Status_{it}} - \boldsymbol{Marital\ Status_{i\cdot}}) + \boldsymbol{\beta_2} \cdot (\boldsymbol{X_{it}} - \boldsymbol{X_{i\cdot}}) + (\varepsilon_{it} - \varepsilon_{i\cdot})$$

For our purposes, the $\boldsymbol{X_{it}}$ term in both the longitudinal specification and in the first specification (from now on called the cross sectional specification) is comprised of a selection of

variables listed in the third column of the table of variables included in section IV, as well as a selection of the marital history variables (years married, years divorced, etc). Many of these control variables were chosen because they are fairly standard in previous literature on the marital wage premium. The main difference in our list of controls and the control variables in other studies is that we are adding spousal wage rates and probability of divorce. The reason for including these variables is to better test the theory of specialization, as has already been discussed in previous sections.

However, as we lack a direct measure for the probability of divorce we will need to estimate a probability of divorce term for each observation in our data set and then include that estimated probability in the econometric specifications given above.

One way to estimate a probability of divorce variable is to use a probit regression. Specifically, we can estimate this variable by the below specification:

3) $$P(Divorce\ In\ Next\ Two\ Years) = \Phi(\alpha_0 + \boldsymbol{\alpha_1} \cdot \boldsymbol{Y_{it}} + \varepsilon_{it})$$

Where $\alpha_0$ is a constant, $\boldsymbol{\alpha_1}$ is a vector of coefficients, and $\boldsymbol{Y_{it}}$ is a vector of personal characteristics. As before, $\varepsilon_{it}$ is the error term. $\Phi(\cdot)$ is the inverse of the cumulative density function for the normal or Gaussian distribution. For our purposes, $\boldsymbol{Y_{it}}$ will be comprised of the variables listed in the fourth column of the table of variables in Section IV. We chose these variables because from an intuitive standpoint it seems that they would likely be correlated with the probability of divorce.

We chose to estimate a probability of divorce over the next two years because the survey rounds of the NLSY in our later period of analysis (2002-2006) are two years apart. The last round of data is 2006, so estimating the probability of divorce over the next two years will allow

us to run the probit model above for the year 2004. To make it clear how we plan to use this probit model, we will estimate a probability of divorce for 2002 and 2004 separately. We will use the coefficients from the 2004 regression to estimate a probability of divorce from our 2006 data. We do this because we lack information on who in our 2006 sample gets divorced in the future.

As said before, after we have predicted a probability of divorce for each person in our analysis, we will use this predicted probability as a variable in the $\boldsymbol{X_{it}}$ term in the cross sectional and longitudinal specifications given at the beginning of this section. In these cross sectional and longitudinal regressions, our reference group, as is typical in the previous literature, will be never married men. As a probability of divorce variable doesn't make much sense for never married men, it will be useful to interact our probability of divorce variable with a dummy variable that is equal to one if an individual is currently married, and equal to zero otherwise. We will also do a third interaction with our cumulative years married variable, for reasons explained in section II.

However, we should mention another method that could allow us to get a rough estimate for the effect divorce probabilities have on the marital wage premium. This would be to leave out our estimates from our probit model and to instead include a dummy in our regressions that is equal to one if a married individual actually did get divorced over the next two years and equal to zero otherwise. We expect actual divorce to be indicative of divorce probabilities, and so this is another method we can use to tease out the effect that probabilities of divorce have on the marital wage premium. However, this also has a drawback. Mainly, we do not know who in our 2006 sample gets divorced in the future, as 2006 is the last round of data in the NLSY. Thus we cannot use this method for analyzing our later time period.

Before we leave our discussion of our probability of divorce variable, we should warn the reader that we expect probability of divorce to be endogenous. Based on discussions on the theory of specialization in section II, we already know that we expect divorce probabilities to have an effect on one's wage; however it also seems probable that wage rates may in part cause divorce probabilities. We also find that spousal wage rates are likely endogenous. If the wage of the spouse has an effect on the husband's wage, it is logical to assume that the wage of the husband has an effect on the wife's wage. While we hope that our probit model will be less prone to endogeneity problems than variables simply indicating whether or not a couple got divorced over a two year time span, we lack a way to completely get rid of these endogeneity problems, and so we should warn the reader that this may be a concern in some of our results. However, we also run regressions without these variables that we can use to draw conclusions about the wage premium. We still feel though, that even with these endogeneity concerns, it is worthwhile to see whether or not our results in regressions including these variables are consistent with specialization or selection. These endogeneity concerns will be discussed again in more detail in section VIII.

In conclusion, these three specifications: the cross sectional specification, the longitudinal specification, and the probit model can explain all the regressions we choose to run in the next section.

We are now ready to present our results in Section VI. The analysis of these results will be presented in Section VII.

## Section VI – Regression Results: Coefficient and Premium Estimates from our Cross Sectional and Longitudinal Regressions

Our results will be presented in tables throughout this section. Detailed analysis of these results will be postponed to Section VII however. To begin with, as some of our cross sectional and longitudinal regressions require us to first estimate a probability of divorce term, we have presented below a table that gives some information about the different probit regressions we ran for this. We ran our probit model to estimate a probability of divorce for each year in our sample, except 2006. The variables used in our probit model can be seen in the table of variables provided in section IV. As we didn't have data on who actually got divorced in the two years following 2006 we had to use the coefficients from the 2004 probit regression to estimate a probability of divorce term for 2006.

We should note here that our probit model is significant at all conventional significance levels in the years 1990-1992, however it has relatively large p-values for the years 2002 and 2004.

PROBABILITY OF DIVORCE ESTIMATION

| | 1990 | 1991 | 1992 | 2002 | 2004 | 2006 |
|---|---|---|---|---|---|---|
| *Mean Probability of Divorce* | 0.0489 | 0.0489 | 0.0588 | 0.0289 | 0.0213 | 0.0242* |
| *Std. Deviation* | 0.0478 | 0.0413 | 0.043 | 0.0225 | 0.0157 | 0.0192* |
| *Prob > chi2* | 0 | 0 | 0.0001 | 0.1566 | 0.6083 | |
| *Pseudo R2* | 0.1047 | 0.0845 | 0.0705 | 0.0654 | 0.056 | |

*estimates for the probability of divorce in the 2006 survey round were gained using the coefficients from the probit model estimating divorce probabilities over the next two years for the 2004 sample.

Part of the difficulty of obtaining precise estimates in the later period may be attributed to the fact that very few people from 2002 through 2006 actually get divorced. Thus we may not have a large enough sample of people who got divorced to get estimates that are significant. That this occurs may be attributed to the age distribution in our different sample periods. From 1990-1992 for example, the average age in our sample is roughly 29.5 years of age, and all ages range from 25 years of age to 35 years of age. However, in our 2002-2006 sample, the average age is roughly 42.5 years of age, and all ages range from 37 years of age to 49 years of age. It makes sense from an intuitive standpoint at least that we would see fewer divorces from people in their mid forties than from people in their late twenties to early thirties.

Now that we have discussed the results from estimating a probability of divorce term for the individuals in our sample, we are ready to start presenting the results from our cross sectional and longitudinal regressions.

The tables are presented in a few pages, and have seven columns, each signifying a different econometric specification used. These columns are the following: *Marital Status, Years Married, Duration of Marriages, Probability of Divorce, Actually Divorced, Spouse's Wage, and Higher Earning Spouse*. As might be guessed, the specification run in the *Marital Status* column includes a full set of control variables (seen in the third column of the table is section V, but not including divorce probabilities or spousal wages), but only includes current marital status dummies to estimate the marital wage premium (no years married terms, divorce probability terms, spousal wage terms, etc).

The specification run in the ***Years Married*** column again includes all the control variables, but this time uses both marital status dummies as well as terms denoting cumulative years married and cumulative years divorced, as well as their squares.

The specification run in the ***Duration of Marriages*** column uses an alternate way to measure marital history. Apart from the full set of control variables and marital status dummies, it includes variables measuring the duration of an individual's first and second marriages, as well as variables measuring the duration of an individual's first and second periods of divorce. Reason's for why me run these regressions that measure marital history in two different ways will be explained later in this section when we start to test the theory of specialization.

In the ***Probability of Divorce*** column we use the same specification as in the *Years Married* column, with the exception that we also include a married individual's estimated probability of divorce (estimated probability of divorce multiplied by an indicator function equal to one if an individual is married) derived partly from our probit regressions, and an interaction term of a married individual's probability of divorce and the cumulative number of years that individual has been married. As hinted earlier, we interact our probability of divorce term with our years married variable, as we think that the years married variable is our best measure of the effect of specialization on wages.

The ***Actually Divorced*** column, replaces the probability of divorce variables in the *Probability of Divorce* column, with a dummy variable that is equal to one if a married individual actually got divorced in the next two years, and equal to zero otherwise. However, for reasons explained on the previous page, we were not able to run this specification for our later sample of individuals.

The specification run in the ***Spouse's Wage*** column is identical to the one run in the *Years Married* column, with the exception that a variable denoting the hourly wage of the spouse is added, as is a variable interacting the wage of one's spouse with the cumulative number of years one has been married. Again, this is done because we are hopeful that the wage of one's spouse will affect the degree of specialization that occurs within a household, and we think that the coefficients on our years married terms are the best indicators of specialization.

Lastly, we have the specification we ran in the ***Higher Earning Spouse*** column. As one might be able to guess by now, this specification is identical to the specification in the *Spouse's Wage* column, except for the fact that we replaced the spousal wage term with a dummy variable that is equal to one if an individual in our sample (i.e. a white male) earns less than their spouse, and equal to zero otherwise. Though the dummy variable used here is obviously a less precise measure of spousal income, it has the benefit of capturing relative wages between a man and his wife, which may influence the type of human capital an individual specializes in more so than spousal wages itself.

Our cross sectional and longitudinal results are now given in the following tables. Just to clarify, we have estimated the effect of marital duration two ways, the effect of divorce probabilities two ways (when possible), and the effect of spousal wages two ways. Also, the key we use to denote the significance of coefficients is the following:

| | | |
|---|---|---|
| Significant at the 1% level | Significant at the 5% level | Significant at the 10% level |

# CROSS SECTIONAL SPECIFICATION: 1990-1992

| | Marital Status | Years Married | Duration of Marriages | Probability of Divorce | Actually Divorced | Spouse's Wage | Higher Earning Spouse |
|---|---|---|---|---|---|---|---|
| *currently married* | 0.0567 | 0.0414 | 0.0573 | -0.0117 | 0.0434 | -0.0100 | 0.1299 |
| *years married* | | 0.0056 | | -0.0014 | 0.0052 | 0.0119 | 0.0029 |
| *years separated/divorced* | | 0.0297 | | -0.0157 | 0.0297 | 0.0311 | 0.0209 |
| *duration of first marriage* | | | -0.0013 | | | | |
| *duration of second marriage* | | | 0.0001 | | | | |
| *Probability of Divorce in Next Two Years* | | | | 1.8488 | | | |
| *(Probability of Divorce) x (Years Married)* | | | | -0.1068 | | | |
| *Divorce in the next two years* | | | | | -0.0395 | | |
| *(Divorce in the next two years ) x (years married)* | | | | | 0.0066 | | |
| *Wage of Spouse* | | | | | | 0.0051 | |
| *(Wage of Spouse) x (Years Married)* | | | | | | -0.0006 | |
| *Higher Earning Spouse* | | | | | | | -0.2166 |
| *(Higher Earning Spouse) x (Years Married)* | | | | | | | 0.0035 |

# CROSS SECTIONAL SPECIFICATION: 2002 – 2006

| | Marital Status | Years Married | Duration of marriages | Probability of Divorce | Actually Divorced | Spouse's Wage | Higher Earning Spouse |
|---|---|---|---|---|---|---|---|
| *currently married* | 0.0805 | -0.0669 | 0.0097 | -0.0378 | **N/A** | -0.1566 | -0.0542 |
| *years married* | | 0.0184 | | 0.0188 | **N/A** | 0.0231 | 0.0200 |
| *years separated/divorced* | | -0.0110 | | -0.0106 | **N/A** | -0.0117 | -0.0110 |
| *duration of first marriage* | | | 0.0046 | | | | |
| *duration of second marriage* | | | 0.0108 | | | | |
| *Probability of Divorce in Next Two Years* | | | | -2.7327 | | | |
| *(Probability of Divorce) x (Years Married)* | | | | 0.1795 | | | |
| *Divorce in the next two years* | | | | | **N/A** | | |
| *(Divorce in the next two years ) x (years married)* | | | | | **N/A** | | |
| *Wage of Spouse* | | | | | | 0.0066 | |
| *(Wage of Spouse) x (Years Married)* | | | | | | -0.0004 | |
| *Higher Earning Spouse* | | | | | | | 0.0494 |
| *(Higher Earning Spouse) x (Years Married)* | | | | | | | -0.0215 |

# LONGITUDINAL SPECIFICATION: 1990-1992

| | Marital Status | Years Married | Duration of Marriages | Probability of Divorce | Actually Divorced | Spouse's Wage | Higher Earning Spouse |
|---|---|---|---|---|---|---|---|
| *currently married* | 0.0114 | -0.0237 | 0.0036 | -0.0367 | -0.0282 | -0.0535 | -0.0177 |
| *years married* | | 0.0181 | | 0.0152 | 0.0188 | 0.0263 | 0.0207 |
| *years separated/divorced* | | -0.0758 | | -0.0757 | -0.0829 | -0.0773 | -0.0772 |
| *Duration of First Marriage* | | | -0.0221 | | | | |
| *Duration of Second Marriage* | | | -0.0220 | | | | |
| *Probability of Divorce in Next Two Years* | | | | 0.4101 | | | |
| *(Probability of Divorce) x (Years Married)* | | | | -0.0240 | | | |
| *Divorce in the next two years* | | | | | 0.0891 | | |
| *(Divorce in the next two years ) x (years married)* | | | | | -0.0178 | | |
| *Wage of Spouse* | | | | | | 0.0028 | |
| *(Wage of Spouse) x (Years Married)* | | | | | | -0.0006491 | |
| *Higher Earning Spouse* | | | | | | | -0.0202 |
| *(Higher Earning Spouse) x (Years Married)* | | | | | | | -0.0097 |

# LONGITUDINAL SPECIFICATION: 2002-2006

| | Marital Status | Years Married | Duration of Marriages | Probability of Divorce | Actually Divorced | Spouse's Wage | Higher Earning Spouse |
|---|---|---|---|---|---|---|---|
| *currently married* | 0.0748 | 0.0673 | 0.0834 | 0.1005645 | | -0.0233916 | 0.0763421 |
| *years married* | | 0.0034 | | -0.0046869 | | 0.00442 | 0.0045562 |
| *years separated/divorced* | | 0.0207 | | 0.0219884 | | 0.0195071 | 0.0206667 |
| *Duration of First Marriage* | | | -0.0076 | | | | |
| *Duration of Second Marriage* | | | 0.0221 | | | | |
| *Probability of Divorce in Next Two Years* | | | | -2.479482 | | | |
| *(Probability of Divorce) x (Years Married)* | | | | 0.124157 | | | |
| *Divorce in the next two years* | | | | | | | |
| *(Divorce in the next two years ) x (years married)* | | | | | | | |
| *Wage of Spouse* | | | | | | 0.0035781 | |
| *(Wage of Spouse) x (Years Married)* | | | | | | -0.0002299 | |
| *Higher Earning Spouse* | | | | | | | 0.0182481 |
| *(Higher Earning Spouse) x (Years Married)* | | | | | | | -0.0125495 |

There are some preliminary observations regarding our cross sectional and longitudinal results that are worth mentioning. For one, we see that we have more significance to our results in our cross sectional regressions than in our longitudinal regressions. We also note that a quick eye-ball glance at these results suggests that our estimates of the various coefficients we are interested in varies (sometimes substantially) depending on whether or not we control for divorce probabilities and spousal wages, and how we control for these variables. This should not be a huge surprise as these variables are more or less endogenous depending on our method of controlling for them. Now we will present some tables which make conclusions about the size of the wage premium in each of our specifications more clear.

Below we have constructed tables showing the predictions for the marital wage premium based on our regression results. These predictions show the predicted difference of wages between a married man and a never married man [ln(married man's wage) – ln(never married man's wage)], assuming that the married man has been married for the average number of years married in our sample, and has never been divorced. These predictions also assume that the married man has no probability of divorce over the next two years, and that his wife's wage is zero. Apart from these aspects, the married man and never married man in our comparison are exactly the same. The prediction tables are given on the next page.

## CROSS SECTIONAL PREDICTIONS: 1990-1992

| | Marital Status | Years Married | Duration of Marriages | Probability of Divorce | Actually Divorced | Spouse's Wage | Higher Earning Spouse |
|---|---|---|---|---|---|---|---|
| *Predicted Premium* | **0.0567** | ***0.0551*** | | ***-0.0189*** | ***0.0555*** | ***0.0259*** | ***0.1321*** |
| *Status Effect* | 0.0567 | 0.0414 | | -0.0117 | 0.0434 | -0.0100 | 0.1299 |
| *Yrs Effect* | 0.0000 | 0.0137 | | -0.0072 | 0.0121 | 0.0359 | 0.0023 |

## CROSS SECTIONAL PREDICTIONS: 2002 – 2006

| | Marital Status | Years Married | Duration of Marriages | Probability of Divorce | Actually Divorced | Spouse's Wage | Higher Earning Spouse |
|---|---|---|---|---|---|---|---|
| *Predicted Premium* | **0.0805** | ***0.1048*** | | ***0.1099*** | | ***0.0802*** | ***0.1597*** |
| *Status Effect* | 0.0805 | -0.0669 | | -0.0378 | | -0.1566 | -0.0542 |
| *Yrs Effect* | 0.0000 | 0.1717 | | 0.1477 | | 0.2368 | 0.2139 |

## LONGITUDINAL PREDICTIONS: 1990-1992

| | Marital Status | Years Married | Duration of Marriages | Probability of Divorce | Actually Divorced | Spouse's Wage | Higher Earning Spouse |
|---|---|---|---|---|---|---|---|
| *Predicted Premium* | **0.0114** | **0.0011** | | **-0.0198** | **0.0001** | **0.0003** | 0.0179 |
| *Status Effect* | 0.0114 | -0.0237 | | -0.0367 | -0.0282 | -0.0535 | -0.0177 |
| *Yrs Effect* | 0.0000 | 0.0249 | | 0.0168 | 0.0283 | 0.0538 | 0.0356 |

## LONGITUDINAL PREDICTIONS: 2002 – 2006

| | Marital Status | Years Married | Duration of Marriages | Probability of Divorce | Actually Divorced | Spouse's Wage | Higher Earning Spouse |
|---|---|---|---|---|---|---|---|
| *Predicted Premium* | **0.0748** | **0.0798** | | **-0.0723** | | **-0.0410** | **0.0792** |
| *Status Effect* | 0.0748 | 0.0673 | | 0.1005645 | | -0.0233916 | 0.0763421 |
| *Yrs Effect* | 0.0000 | 0.0124 | | -0.1728 | | -0.0176 | 0.0028 |

Before we analyze these results in the next section, we have enough here to make some preliminary observations. It seems that the wage premium has grown over time. We should keep in mind that our later analysis is focused on the same people in our early analysis, only we look at them ten to fifteen years later. While we have included age as a control variable in all our regressions, there is still some concern that this trend in the premium is due the fact that the individuals are older in 2002-2006 than in 1990-1992. This concern will be discussed in more detail in Section VIII.

Also, we note simply from our prediction tables that our estimate of the wage premium from 1990-1992 is lower than is estimated in some previous literature focusing on this time period, specifically Gray's paper from 1997. This may be attributed to the fact that his analysis of the premium used a mixture of younger and older individual's, which we think would tend to place upward pressure on the estimate of the wage premium. However, as will be seen in the upcoming section, our interpretation of the main causes of the premium during 1990-1992 is similar to Gray's.

This concludes the presentation of the main results from our regressions. We will now move on to the next section, which is devoted to using these results to answer the fundamental questions of this paper (though not necessarily in order). Mainly, how has the marital wage premium changed over time? And also, what economic theories that explain the marital wage premium are most consistent with our data?

## Section VII – Using our Results to Test the Theory of Specialization and the Selection Hypothesis

This section will be divided into three parts. The first part will use the results from our regressions using data from (1990-1992) in order to test which theories best explain the wage premium during this early time period. The second part will use almost identical techniques to test which theories best explain the marital wage premium in the later time period (2002 – 2006). Lastly, the third section will combine our conclusions from the first two sections in order to make some conclusions about how the marital wage premium has changed over time.

### Part I – The Early Period Analysis

As stated before, we are interested mainly in testing the theory of specialization and the selection hypothesis. There are many ways we can use our results from the previous section to do this.

In fact, there are six main things we will look at to see how well the theories of specialization and selection explain our data.

1) **We will compare our results with the prediction tables created in section II of this paper and see if the signs on our coefficients are as expected**

2) **We will compare our results from our cross sectional and longitudinal regressions. If our results are very different, than we may expect that fixed unobserved characteristics are driving the wage premium, which suggests the selection hypothesis is more important than the theory of specialization in determining the marital wage premium.**

3) **We will compare the relative importance of cumulative years married and current marital status. If a large majority of the predicted wage premium is due to a cumulative years married effect, than this suggests that the theory of specialization is driving our results.**

4) **We will also look at the specification including the duration of an individual's first and second marriage rather than their cumulative years married variable. We have reason to believe (as will be explained later) from the theory of specialization that the duration on an individual's first and second marriage should have different coefficients. Thus we can test for whether or not these coefficients are different to test the theory of specialization.**

5) **We will look at how controlling for the probability of divorce effects our estimates of the wage premium. The theory of specialization has clear predictions regarding the effect high probabilities of divorce should have on the wage premium. By seeing if these predictions fit our data we can test whether or not our data is consistent with the theory of specialization.**

6) **Lastly, we will look at the effect that controlling for spousal wages has on the wage premium. Again, the theory of specialization has clear predictions for how spousal income should effect the marital wage premium, so by checking whether or not our data is consistent with these predictions, we can further test for the relevance of the theory of specialization in explaining the wage premium.**

1. First, we can see how our results compare to the predictions made from the theory of specialization. These predictions can be found in table 2 of section II of this paper. This information is presented in the table below.

For the purposes of creating this table we used the results from our early longitudinal regressions. We do this because the longitudinal specification should give a more accurate measure of the marital wage premium as it allows us to control for unobserved fixed effects.

As can be seen, our data is mostly consistent with the specialization predictions made in section II. The one exception is that we find negative coefficients on our currently married dummy in our early longitudinal regressions. However, this does not necessarily contradict the theory of specialization. For one, these negative coefficients are not statistically significant, and at worst they simply mean that other forces besides the theory of specialization are at least somewhat influencing the wage premium.

| Explanatory Variables | Sign of Variable as Predicted by the Theory of Specialization* | Sign of Variable as Found in our Data(Early Longitudinal Regressions) |
|---|---|---|
| Currently Married Dummy | Should Have Little to no Effect on the Wage Premium | Mostly Negative |
| Years Married | Positive | Positive |
| Years Divorced | Negative | Negative |
| (Probability of Divorce) x (Years Married) | Negative | Negative |
| (Wage of Spouse) x (Years Married) | Negative | Negative |

**predictions made in Table 2 of section II of this paper*

However, we would not expect to see these negative coefficients from the selection hypothesis, as that theory clearly predicts that the coefficient on our currently married dummy should be positive.

Thus after comparing our results with the predictions made in table 2 of section II, we can say that our data is consistent with the theory of specialization driving the wage premium that exists in our longitudinal regressions (after we control for fixed effects). Now we can compare our results to table 3 in section II, which gives predictions based on the selection hypothesis. Two tables are provided below. One uses results from our longitudinal specification, and one from our cross sectional specification.

| Explanatory Variables: | Sign of Variable as Predicted by the Selection Hypothesis* | Sign of Variable as Found in our Data(Early Longitudinal Regressions) |
|---|---|---|
| Currently Married Dummy | Positive | Mostly Negative |
| Years Married | Should Have Little to no Effect on the Wage Premium | Positive |
| Years Divorced | Should Have Little to no Effect on the Wage Premium | Negative |
| (Probability of Divorce) x (Years Married) | Should Have Little to no Effect on the Wage Premium | Negative |
| (Wage of Spouse) x (Years Married) | Should Have Little to no Effect on the Wage Premium | Negative |

**predictions made in Table 3 of section II of this paper*

Here we find that the selection hypothesis does not do an adequate job of explaining our early longitudinal results. However, it does a better job of explaining our cross sectional results, which are generally marked with large positive coefficients on our currently married dummy and exhibit small coefficients on our cumulative years married variable. However, we still see evidence that the selection hypothesis does not completely explain all of our cross sectional results.

| Explanatory Variables: | Sign of Variable as Predicted by the Selection Hypothesis* | Sign of Variable as Found in our Data(Early Cross Sectional Regressions) |
|---|---|---|
| Currently Married Dummy | Positive | Usually Positive and Large |
| Years Married | Should Have Little Effect on the Wage Premium | positive but very small |
| Years Divorced | Should Have Little Effect on the Wage Premium | generally positive |
| (Probability of Divorce) x (Years Married) | Should Have Little Effect on the Wage Premium | Negative |
| (Wage of Spouse) x (Years Married) | Should Have Little Effect on the Wage Premium | very small |

**predictions made in Table 3 of section II of this paper*

2. Below (on the next page) is a table showing the predictions made for the marital wage premium from our cross sectional and longitudinal regressions. We see that the estimates of the marital wage premium from our longitudinal regressions are generally much smaller than the estimates from our cross sectional regressions. This suggests that unobserved characteristics are a very important determinant of wages for married men. Indeed, after controlling for these characteristics, our results suggest that the wage premium virtually disappears (ranging from roughly negative -2% to 2%, with most estimates closer to 0.01% – 0.1%)

Combined with what we know from comparing our longitudinal results with our predictions from table 2 of section II, we can so far conclude that unobserved personal characteristics are the driving force behind the marital wage premium in this period, which supports selection as the driving force behind the wage premium. However, once you control for

these fixed effects as best you can, the remaining premium, though miniscule, is fairly well explained by the theory of specialization.

As a note to the reader, attributing a large majority of the premium to selection and a small part to specialization is consistent with previous literature focusing on the wage premium in the late 80's to early 90's (Gray,97).

| | Marital Status | Years Married | Probability of Divorce | Actually Divorced | Spouse wage | Higher Earning Spouse |
|---|---|---|---|---|---|---|
| *Prediction from Cross Section Regressions(90-92)** | 0.0567 | 0.0551 | -0.0189 | 0.0555 | 0.0259 | 0.1321 |
| *Prediction from Longitudinal Regressions(90-92)** | 0.0114 | 0.0011 | -0.0198 | 0.0001 | 0.0003 | 0.0179 |

**A quick reminder of how predictions were created. They are the predicted difference in the natural log of hourly wage between a never married man and a currently married man who has been married for the average number of years married in our sample, but who is otherwise identical to the never married man (i.e. no probability of divorce & no spousal wages among other things)*

3. Next we present a table which lets us see the relative effects of marital status and cumulative years married on the wage premium. Again, this table was constructed using results from our early longitudinal regressions, for reasons explained earlier.

The table below simply presents the amount of the predicted wage premium given in the bottom row of the table above that is comprised of a marital status effect, and the amount of the predicted premium that is comprised of a years married effect.

| | Marital Status | Years Married | Probability of Divorce | Actually Divorced | Spouse wage | Higher Earning Spouse |
|---|---|---|---|---|---|---|
| *Status Effect* | 0.0114 | -0.0237 | -0.0367 | -0.0282 | -0.0535 | -0.0177 |
| *Yrs Effect* | 0 | 0.0249 | 0.0168 | 0.0283 | 0.0538 | 0.0356 |

We see here that in all longitudinal specifications in which we include a cumulative years married term, we find that cumulative years married is far more important in so far as increasing wages goes than is marital status. This is predicted by the theory of specialization, and suggests that the marital wage premium that remains after controlling for fixed effects is largely explained by specialization in human capital. This is in line with our earlier results, which find the majority of the wage premium being explained by fixed effects, but the remaining premium being well explained by specialization.

4. Now we will look at our longitudinal specification which uses the duration of an individual's separate periods of marriage and divorce as a substitute for the cumulative years married and divorced terms in our other regressions.

We do this because we suspect that the durations of an individual's first and second marriages may have different coefficients under the theory of specialization. This is because the rate of human capital accumulation is likely not constant over time. Thus a certain amount of time spent acquiring market related human capital early in a marriage is likely to be more or less productive than an equal amount of time spent acquiring the same type of capital later in the marriage. For this reason, we might also expect that the years in an individual's first marriage are

more or less productive in terms of acquiring new human capital than the years in an individual's second marriage.

The table below provides some relevant information regarding the results of our early longitudinal regression that uses this alternative measure for an individual's marital history.

These results are quite surprising. For one, we see that the coefficients on the durations of an individual's first and second marriage are negative. This is especially confusing given the fact that our cumulative years married variable can be seen as the sum of these two duration variables, and that regressions using our cumulative years married variable yield a positive coefficient for this variable.

| | **Duration of Marriages** |
|---|---|
| *Duration of 1st Marriage(Coeff)* | -0.0221 |
| *Duration of 2nd Marriage(Coeff)* | -0.0220 |
| *Duration of 1st Marriage - Duration of 2nd Marriage* | -0.0001 |
| *Prob > F* | 0.9968 |

We aren't sure then how to explain why the coefficients in the above table should be negative. We also see from the bottom row of the table the results from an F test testing the hypothesis that the coefficients on the two marital duration variables are the same. We conclude that the coefficients are not different from each other in a statistically significant manner. Both of

these facts qualify our earlier conclusion that the theory of specialization completely explains the premium existing in our longitudinal analysis.

5. Below we finally examine the effect that controlling for the probability of divorce has on the marital wage premium. To remind the reader of what we did in the previous section, we estimated the probability of divorce in our early period in two ways. One was to use the results from our probit model, and the other was to simply include a dummy variable equal to one if a married individual actually did get divorced sometime in the next two years and equal to zero otherwise. We have included results from both ways of controlling for divorce probabilities in the table provided. Again, these results come from our early longitudinal regressions.

| | **Probability of Divorce** | **Actually Divorced** |
|---|---|---|
| *Probability of Divorce in Next Two Years* | 0.4101 | |
| *(Probability of Divorce) x (Years Married)* | -0.024 | |
| *Divorce in the next two years* | | 0.0891 |
| *(Divorce in the next two years ) x (years married)* | | -0.0178 |
| *Prediction of Premium Controlling for Divorce* | -0.0198 | 0.0001 |
| *Status Effect* | -0.0367 | -0.0282 |
| *Yrs Effect* | 0.0169 | 0.0283 |

This table has some results we expected but also some surprises. For one, we notice that the intercept effect of the probability of divorce (seen by the coefficients in the first and third

rows) is surprisingly large and positive. This runs counter to what we would expect from the selection hypothesis. However, we also note that for our prediction of the wage premium in this table, as in all our predictions, we assumed the individual had zero probability of divorce. The positive intercept effect of divorce probabilities on the wage premium may then explain why the status effect in the table above is negative. In real life, we would think that married people have a nonzero probability of divorce, and so as our probability of divorce term only applies to married people in our regression, the status effect for most of our sample is not so negative, as they will also benefit from the positive coefficients on our probability of divorce terms.

However, this discussion is somewhat beside the point when it comes to looking at how our results here support or don't support the theory of specialization. The theory of specialization implies that the marital wage premium accrues over time as people specialize investments in human capital. Thus to know the effect of divorce probabilities on the marital wage premium, at least according to the theory of specialization, we need to look at how these probabilities effect changes in human capital investment patterns. From our regressions, the coefficient on the cumulative years married term is the best measure of how intensive specialization in market related human capital is. Thus interacting the probability of divorce with cumulative years married will give us our best measure for how divorce probabilities affect human capital investment patterns. According to the theory of specialization, we should see less specialization in market related human capital as probabilities of divorce increase. This means that high probabilities of divorce should lead to a slower growth in wages. For our data, this is equivalent to saying that the coefficient on the term interacting divorce probabilities and cumulative years married should be negative. Indeed, looking at our table confirms that this coefficient is indeed negative, regardless of how we measure the probability of divorce.

Thus our results from controlling for the probability of divorce are consistent with the theory of specialization. Again, however, we see that after controlling for fixed effects, the effect of marital status is no longer necessarily positive. As this is not explained by the selection hypothesis, we suspect that getting rid of fixed effects largely gets rid of the positive effect that selection has on wages. Ultimately these results are consistent with the previous notion we had that whatever premium exists after removing fixed effects can be attributed to specialization.

6. Finally we will look at the effect of controlling for spousal wages on the marital wage premium. Again, we controlled for the wages of an individual's spouse in two ways. One was to simply include the spouse's wage into our regression, and the other was to include a dummy variable equal to one if an individual has a higher earning spouse, and zero otherwise. The table on the top of the next page presents some of our more important results from our early longitudinal regressions that focused on controlling four spousal wages.

We see here that after controlling for fixed unobservable characteristics, that the intercept effect of having a high earning spouse is mixed. This can be seen by the fact that we have a positive coefficient in the first row of the table below, and a negative coefficient in the third row. We note further than the positive coefficient in the first row of the table below is extremely small. These facts signal that after controlling for fixed effects, the selection hypothesis no longer does an adequate job of explaining the remaining wage premium.

However, we see from the coefficients on our interaction terms that higher wages for one's spouse is correlated with lower wages for oneself. Again this effect is rather small, but nonetheless these negative coefficients are what we would expect from the theory of specialization. This is because as stated above, the coefficients on our years married terms are

the best measures for the degree to which an individual is specializing in market related human capital, and so the coefficients we get when we interact our measures of spousal wages with our cumulative years married variable indicate how increases in spousal wages effect the degree of household specialization.

| | Spouse wage | Higher Earning Spouse |
|---|---|---|
| *Wage of Spouse* | 0.0028426 | |
| *(Wage of Spouse) x (Years Married)* | -0.0006491 | |
| *Higher Earning Spouse* | | -0.0202 |
| *(Higher Earning Spouse) x (Years Married)* | | -0.0097 |
| *Predicted Premium Controlling for SP wage* | 0.0003 | 0.0179 |
| *Status Effect* | -0.0535 | -0.0177 |
| *Yrs Effect* | 0.0538 | 0.0356 |

The negative coefficients we see suggest that an individual slightly decreases their investments in market related human capital as their spouse's wage increases. As we think that there is less incentive for the husband to specialize in market related human capital as the value of their spouse's time in the labor market becomes high, this result is consistent with the theory of specialization.

Overall, we again see that our results are consistent with the theory of specialization. As the intercept effect of being married in these regressions is actually negative, we see that the small premium that remains after controlling for fixed effects can currently only be explained by the theory of specialization.

In conclusion, our early data has cross sectional estimates of the premium ranging from over 2% to over 13%, with most estimates falling around 5% or 6%. However, when we control for fixed effects we find that the wage premium virtually disappears in most of our longitudinal regressions. However, we find that the little bit of premium that remains is generally well explained by the theory of specialization, which was highly consistent with 5 out of the 6 methods we used to test the relevance of that theory. The one method which gave mixed signals regarding the theory of specialization was when we looked at our results from the specification using our alternate measure of marital history. However, as explained earlier, there are some questions regarding the validity of those results that we currently don't know how to answer. This fact, along with the fact that all our other methods supported the theory of specialization so strongly, causes us to conclude that the theory of specialization is fairly well supported by our data.

In fact, even though the estimate of the marital premium after we control for fixed effects is small, we see that often times the effect of cumulative years married is not. Rather this effect can be somewhat large but is often counteracted by negative coefficients on our marital status terms. This may signal that selection is still present in our longitudinal results, but because it is acting in the opposite direction to what we expect it is not contributing to the marital wage premium in our longitudinal regressions, thus meaning that whatever premium exists is attributable to the theory of specialization.

But why should we think that selection into marriage should be correlated with lower wages in our longitudinal results? We don't have a clear answer for this. One idea might be that many of the characteristics that are valued in both the labor and marriage markets are generally fixed. However, perhaps not all are, and someone could still be selected into marriage partly based on the fact that he has a characteristic at the time of marriage, but if this personal characteristic is not fixed, then over time it is possible that such a characteristic could change until it was no longer correlated with higher wages. However, ultimately we can't test this idea from the regressions we ran, and there may be other, better explanations not involving the selection hypothesis.

In general though, we see that the majority of the wage premium is determined by fixed unobservable characteristics. This suggests that the selection hypothesis drives the majority of the wage premium. However, after controlling for these fixed effects a small premium still remains that is attributed to specialization.

The next obvious question is how do these results compare with our results from our later data?

## Part II – The Later Period Analysis

We can use the same techniques that we used when analyzing our early data to analyze what theories best explain the marital wage premium in our later data.

1. The first thing to do is to take a quick look at how the results from our later data compare with the tables of predictions (tables 2 and 3) in section II. First we look at the predictions we made from the theory of specialization. A table comparing these predictions to our results is given on the next page.

Again, like we did when analyzing our early data, our results come from our later longitudinal regressions. The reasons for doing this are the same as when we analyzed our early data. We see that unlike the early period, our longitudinal results this period are not in general consistent with the predictions from the theory of specialization. The signs of several coefficients are opposite of what the theory of specialization predicts. Specifically, the coefficient on our cumulative years divorced variable and the coefficient on our term interacting an estimated probability of divorce and a cumulative years marred variable are both positive, when the theory of specialization clearly predicts that they should be negative. We also see positive coefficients on our cumulative years divorced term.

| **Explanatory Variables** | **Sign of Variable as Predicted by the Theory of Specialization*** | **Sign of Variable as Found in our Data(Later Longitudinal Regressions)** |
|---|---|---|
| Currently Married Dummy | Close to Zero | Positive |
| Years Married | Positive | Usually Positive & Small |
| Years Divorced | Negative | Positive |
| (Probability of Divorce) x (Years Married) | Negative | Positive |
| (Wage of Spouse) x (Years Married) | Negative | Negative |

**predictions made in Table 2 of section II of this paper*

So unlike in our early data, we find that the theory of specialization up to this point does a relatively poor job of explaining our results. Now we will look at how our results compare to the predictions made from the selection hypothesis.

We see from the table below, which again uses results from our later longitudinal regressions, that while our results are consistent with the theory of selection, we also see that selection may not be the only force influencing the wage premium. We see that the coefficient on our currently married dummy is positive as predicted. The selection hypothesis however doesn't give any particular reason to think there should be much of an effect from out other variables. While some of the coefficients on the variables in the below table are miniscule, other such as the coefficient on our cumulative years separated/divorced variable are larger in magnitude. That is why while our results do not contradict the selection hypothesis; they suggest that the selection hypothesis may not be able to explain all our results.

| **Explanatory Variables:** | **Sign of Variable as Predicted by the Selection Hypothesis** | **Sign of Variable as Found in our Data(Later Longitudinal Regressions)** |
|---|---|---|
| Currently Married Dummy | Positive | Positive |
| Years Married | Should Have Little to no Effect on the Wage Premium | Usually Positive & Small |
| Years Divorced | Should Have Little to no Effect on the Wage Premium | Positive |
| (Probability of Divorce) x (Years Married) | Should Have Little to no Effect on the Wage Premium | Positive |
| (Wage of Spouse) x (Years Married) | Should Have Little to no Effect on the Wage Premium | Negative |

**predictions made in Table 3 of section II of this paper*

We also note from the table below that the theory of selection seems to explain our later cross sectional results about as well as our later longitudinal results. Again, before controlling for the part of the selection hypothesis attributed to fixed effects, we still see that while our results are not inconsistent with the selection hypothesis, they are not explained in whole by that theory.

| Explanatory Variables: | Sign of Variable as Predicted by the Selection Hypothesis | Sign of Variable as Found in our Data(Later Cross Sectional Regressions) |
|---|---|---|
| Currently Married Dummy | Positive | Positive |
| Years Married | Should Have Little to no Effect on the Wage Premium | Usually Positive & Small |
| Years Divorced | Should Have Little to no Effect on the Wage Premium | Positive |
| (Probability of Divorce) x (Years Married) | Should Have Little to no Effect on the Wage Premium | Positive |
| (Wage of Spouse) x (Years Married) | Should Have Little to no Effect on the Wage Premium | Negative |

**predictions made in Table 3 of section II of this paper*

2. Now we will compare our results from our later cross sectional and later longitudinal regressions. Below is the predicted wage premium for each of our later regressions in the table below. Again, if we find that our predicted premium is very different in these two specifications, we suspect that fixed unobserved personal characteristics are driving the wage premium, which would tend to support the selection hypothesis.

| | Marital Status | Years Married | Probability of Divorce | Spouse wage | Higher Earning Spouse |
|---|---|---|---|---|---|
| *Prediction from Cross Sectional Regressions(2002 - 2006)* | 0.0805 | *0.1048* | *0.1099* | *0.0802* | *0.1597* |
| *Prediction from Longitudinal Regressions(2002 - 2006)* | 0.0748 | 0.0798 | -0.0723 | -0.041 | 0.0792 |

We see that the difference in our predicted premium depends specifically on what specification we are running. For instance, before we control for divorce probabilities or spousal wages, we see that fixed effects cause the wage premium to rise roughly 1% to 3% (the difference between the predicted premium in our marital status and years married specification). However, when we control for the probability of divorce and spousal wages, we see that fixed effect have a huge effect on our predicted premium. Thus our results are here somewhat mixed as far as the selection hypothesis goes. Our data suggests that fixed effects are somewhat important, and so we would also think that the selection hypothesis is useful in explaining our results. However, we can't say for certain that fixed effects are driving our results to the same extent as they were in the period 1990-1992.

This is again consistent with our previous findings that the selection hypothesis is consistent with our data, however it is not clear by simply looking at the effect that fixed effects have on our regressions how much of the wage premium is actually due to the selection hypothesis.

3. Next we will look at the relative importance of marital status and marital history in determining the wage premium. For the table presented below we use the results from our later longitudinal regressions. We can clearly see the amount of the predicted wage premium that is attributable to the effect of current marital status and the effect of cumulative years married. Again, if marital status is much more important than cumulative years married, that is consistent with the selection hypothesis but is evidence against the theory of specialization.

| | Marital Status | Years Married | Probability of Divorce | Spouse wage | Higher Earning Spouse |
|---|---|---|---|---|---|
| *Status Effect* | 0.0748 | 0.0673 | 0.1006 | -0.0234 | 0.0763 |
| *Yrs Effect* | 0.0000 | 0.0124 | -0.1728 | -0.0176 | 0.0028 |

We see from this table, that unlike in our early period, most of the wage premium remaining after we control for fixed effects is attributable to one's current marital status, instead of to one's marital history. While the relatively large and positive status effects support the selection hypothesis, the always low and sometimes negative years effect is evidence against the theory of specialization. There is one regression where our status effect is negative, however the fact that our predictions assume zero spousal wages may again explain this result. As the majority of people in our sample likely have spouses with non-zero wages, we may see a different and even positive status effect for most of the people in our sample. This along with the fact that all our other regressions estimate a very large status effect suggests that these results are fairly well explained by the selection hypothesis.

To sum things up, we see that the selection hypothesis can be used to explain most the premium that remains after we control for fixed effect. There is again little evidence for specialization.

4. Now we again look at our specification that uses the durations of an individual's separate marriages rather than their cumulative years married. As explained when analyzing our early data, we suspect from the theory of specialization that the duration of an individual's first marriage may have a different coefficient than the duration of an individual's second marriage. The table below presents some key results from this regression. Again, the results are from our later longitudinal regressions.

These results are a bit hard to interpret. On one hand, we see from the last row of this table that the coefficients on the durations of marriage are significantly different from each other (in a statistical sense) at a 10% significance level. This is not something that would necessarily be predicted from the selection hypothesis. However, the sign of the duration of the first marriage is negative and extremely small, which seems to be evidence against the theory of specialization.

Ultimately these results are not adequately explained by either the theory of selection of the theory of specialization.

| | Duration of Marriages |
|---|---|
| *Duration 1st Marriage* | -0.0076 |
| *Duration of 2nd Marriage* | 0.0221 |
| *(Duration of 1st Marriage) - (Duration of 2nd Marriage)* | -0.0297 |
| *Prob > F* | 0.0762 |

5. Next we will look at the effect that controlling for divorce probabilities has on the wage premium. A table similar to the one presented when analyzing the early period of our data is given on the next page.

What we see here is more evidence against the theory of specialization. We see this mainly because the coefficient on our term interacting divorce probabilities is not the sign we would expect from the theory of specialization. That theory predicts that higher probabilities of divorce will lower incentives to specialize in market related human capital, and thus slow down the growth of wages. However, the positive coefficient on our interaction term in the table below says that the opposite is true: higher probabilities of divorce are correlated with an increase if the growth of wages. This further supports the evidence gained so far that the theory of specialization does a poor job explaining our data from this period

| | Probability of Divorce |
|---|---|
| *Probability of Divorce in Next Two Years* | -2.4795 |
| *(Probability of Divorce) x (Years Married)* | 0.1242 |
| *Prediction of Premium Controlling for Divorce* | -0.0723 |
| *Status Effect* | 0.1006 |
| *Yrs Effect* | -0.1728 |

We also see that there is a negative coefficient on the term that gives a probability of divorce for married men (seen in the first row). This is overall consistent with the selection hypothesis. We would suspect that people with high probabilities of divorce may have been originally selected into marriage by "mistake". That is, their partner may have believed they had certain characteristics valued in both the marriage and labor market, when in fact they didn't. Thus we would expect from the selection hypothesis a negative effect of high divorce probabilities on the coefficient for currently being married. This is exactly what we get with the negative coefficient in the first row of the table above.

Again, we also see as we saw earlier, that the marital status effect on the predicted wage premium is large and positive, while the years married effect is negative and even larger in magnitude.

Overall these results are consistent with the idea that the selection hypothesis is an important determinant of the wage premium during this time, but that the theory of specialization

is not. However, as the selection hypothesis alone would seem to predict a high premium (based on the large status effect in the above table), and the theory of specialization does not predict a negative years married effect; we also have to conclude that the selection hypothesis alone cannot explain all of these results.

6. Finally we will look at the effect that controlling for spousal wages has on the marital wage premium. Again, a table similar to the one presented during our analysis of our early data is present on the next page.

First note that in this table we actually see some evidence that is consistent with the theory of specialization. Mainly, we see negative coefficients on our terms interacting spousal wage and cumulative years married. This means that higher spousal wages are correlated with slower growth in wages, which is what the theory of specialization predicts, as there is again less incentive to specialize in market related capital when one's spouse is a very high earner.

However we note that while the effect on the predicted wage premium of cumulative years married is positive in the second column in our table, it is negative in the first column. This negative number is something that is not consistent with the theory of specialization. To remind the reader, the difference between the first and second columns of the table is that in the first column we include actual measurements of spousal wages, while in the second column we include only a dummy variable that tells us whether or not an individual has a higher earning spouse.

| | Spouse wage | Higher Earning Spouse |
|---|---|---|
| *Wage of Spouse* | 0.0036 | |
| *(Wage of Spouse) x (Years Married)* | -0.0002 | |
| *Higher Earning Spouse* | | 0.0182 |
| *(Higher Earning Spouse) x (Years Married)* | | -0.0125 |
| *Predicted Premium Controlling for SP wage* | -0.0410 | 0.0792 |
| *Status Effect* | -0.0234 | 0.0763 |
| *Yrs Effect* | -0.0176 | 0.0028 |

It's not clear how to interpret what these findings say about the selection hypothesis. The positive coefficients on our spousal wage variable (which only applies to married men) and our higher earning spouse dummy (again only applying to married men) indicate that higher earning spouses are correlated with larger benefits being attributed simply to one's current marital status.

It's not clear if this consistent or not with the selection hypothesis. This is because we don't know enough about the selection process. Do women with high wage rates select for some characteristics in men to a lesser or greater degree than women with low wage rates? The intuitive answer is yes, however we can't be sure, and we don't know without further analysis how the selection process is supposed to be different for women with high and low wage rates.

The negative status effect in the first column of the table above might also appear to be evidence against the selection hypothesis. However, we note that our prediction of the premium assumed that our married individual had a wife with no wage rate. Given this probably doesn't hold for the vast majority of people, and the fact that we have a positive coefficient for our spousal wage variable, it is possible that for most people, the status effect would actually be positive. We also cannot preclude the possibility that being selected into marriage with a wife who has no earning potential is correlated with lower wages. This would go some way in explaining our negative status effect in the first column, and goes back to the idea that women with different earning potentials may have different selection criteria for men, and thus men who get selected into marriage with low earners may have different characteristics than men who get selected into marriage with high earners. For these reasons, we will refrain from making conclusions regarding the selection hypothesis from this data.

To sum things up, we find that controlling for spousal wages yields some results that are consistent with the theory of specialization but yields other results that contradict the theory of specialization. We can make no conclusions regarding the selection hypothesis from this data.

This concludes our analysis of our 2002 – 2006 data. We found very little reason to think that specialization was an important determinant of wages. We did find evidence however that the selection hypothesis can explain a large amount of our later results. However, we also found some results that could not adequately be explained by either theory.

We are now ready to move on to the last part of our data analysis, which will compare the results from our early and later periods in order to make some conclusions about how the marital

wage premium has changed. Much of the analysis will follow fairly directly from what we have already discussed.

## Part III – How has the Marital Wage Premium Changed?

Below is a table that presents the predicted values of the wage premium for the years 1990-1992 and the years 2002-2006. Here we are using results from our longitudinal regressions.

| | Marital Status | Years Married | Probability of Divorce | Spouse wage | Higher Earning Spouse |
|---|---|---|---|---|---|
| *Predictions from Early Longitudinal Regressions* | 0.0114 | 0.0011 | -0.0198 | 0.0003 | 0.0179 |
| *Predictions from Later Longitudinal Regressions* | 0.0748 | 0.0798 | -0.0723 | -0.041 | 0.0792 |

We see that estimates for the overall marital wage premium tend to be higher in our later period. We may have gotten a more similar estimate (after controlling for fixed effects) for the wage premium in our regressions where we controlled for the probability of divorce had we assumed an individual had an average probability of divorce rather than no probability of divorce. Likewise, we may have gotten a larger and positive prediction for the premium in our regression where we included spousal wage values had we assumed that an individual had a spouse with an average wage rate rather than a spouse whose wage rate was zero.

Overall it seems that after controlling for fixed effects the predicted wage premium in the early period is no more than 2%, while the wage premium in the later period is as large as 7% or 8%.

Before controlling for fixed effects we find the following predictions for the marital wage premium:

| | Marital Status | Years Married | Probability of Divorce | Spouse wage | Higher Earning Spouse |
|---|---|---|---|---|---|
| *Predictions from Early Cross Sectional Regressions* | 0.0567 | 0.0551 | -0.0189 | 0.0259 | 0.1321 |
| *Predictions from Later Cross Sectional Regressions* | 0.0805 | 0.1048 | 0.1099 | 0.0802 | 0.1597 |

Again we see that before controlling for fixed effects, estimates of the premium are always larger in our later period than in our earlier period. We conclude with saying that the majority of our evidence supports a wage premium that has been growing since the early 90's.

However, we need to make it clear that we are comparing the wage premium experienced by a group of men in the early 90's with the wage premium experienced by that same group more than a decade later. Thus, as will be discussed in the next section, this rising wage premium we found may be due to the fact that we are looking at a different distribution of ages in our early and later analysis. The fact that our predictions assumed no time had been spent divorced or separated may allow for some upward pressure in the wage premium as people get older. While we clearly would predict an increasing premium from the theory of specialization, our results from the 2002-2006 data don't support this conclusion. However, we can also explain an increasing wage premium with age with the selection hypothesis. Better marriage partners we think would have higher wages due to selection. The longer someone is married, without being

divorced, would help verify the fact that an individual may have characteristics that are valued in both the job and marriage markets. Thus it makes sense the older people get (without getting divorced, as our prediction assume), the more upward pressure selection would have on the wage premium.

As far as the causes behind the wage premium, we find fairly strong evidence that some specialization is occurring in our early period, though we still find that the majority of the premium in the early period can be attributed to fixed effects. The premium that remains after controlling for fixed effects in the early period can be attributed to specialization however.

We also see that while our later data is consistent with the selection hypothesis and that selection can explain much of the wage premium from that time, the selection hypothesis alone cannot explain all of our results from that period. We find however, that from 2002-2006, the vast majority of our evidence contradicts the theory of specialization

Our next section will be devoted to discussing some possible econometric concerns and whether or not they pose a significant problem.

## Section VIII – Econometric Concerns: Are our Results Valid

We can group concerns with our results into three main areas. While we believe that many of these concerns can be alleviated, not all of them can. The three main areas of concern are:

1) The form of our econometric specification
2) Endogeneity concerns among two important variables in our analysis: probability of divorce and spousal wages.

3) Endogeneity concerns for other control variables used in our study

First we will discuss whether or not our econometric specification had the correct form. This mainly deals with things like whether or not we interacted and squared all the terms we should have. Our main strategy in creating the form of our specifications was to try to create a specification which would allow us to infer certain things about the theory of specialization and the selection hypothesis, not one which would capture the causation of wage rates with the utmost detail. As the coefficient on marital status was the main tool we had to make inferences regarding selection, and the coefficient on years married was the main tool we had to make inferences regarding specialization, the interaction terms we decided to include in our regression always were interactions with either cumulative years married or a currently married dummy variable. We felt that getting away from focusing on these variables would not be productive as for the most part it was with these variables alone that we could make inferences regarding selection and specialization. Because of this we are fairly confident that other interactions not including these variables (which were interactions left out of our regressions) would not have dramatically changed the interpretation of our results.

As for the terms we decided to interact with marital status and cumulative years married, these were variables that we felt would have an important impact on either the selection process or the degree of specialization within a household. Depending on what we were controlling for at the time, these variables would have either been divorce probabilities or indicators, or spousal wage variables. We didn't feel that the other variables in our regression would have had a predictable impact on selection or specialization, and so we didn't interact those variables with marital status or cumulative years married.

One important question that should be dealt with however is whether or not we adequately controlled for age. As age is the primary difference between our earlier sample and older sample, it seems that concerns about age may cause us to question the upward trend in the wage premium we saw in our data since the early 90's. Perhaps this upward trend exists because we didn't do an adequate job of controlling for age. This concern is not entirely invalid. In our regressions we controlled for age by simply including a term that specified an individual's current age. Is it possible that the characteristics an individual gets selected into marriage for depend on the age of the individual and the age of his potential spouse? Is it possible that older individuals will have unique incentives to either specialize less or more than their younger counterparts? Neither of these questions can be answered firmly in the negative, and so one can argue that a more realistic specification would have interacted age with our marital status and or cumulative years married variables. It may indeed be the case that we see a different trend in the wage premium over time if we did a more complete job of controlling for age.

However, we feel justified for the purposes of this paper in the specifications we used. Clearly some compromises have to be made between the specification used and the way we think the world works in greater detail. We chose to choose a more simple method for controlling for age simply because we felt that as we were already interacting marital status and cumulative years married with other variables, interpretation of our results would become too messy if we also included a more advanced method of controlling for age. Nonetheless, the reader should be aware that the difference in ages between our early and later sample is very large, and thus may be influencing the upward trend in the wage premium that we saw in section VII.

Now we will move on to a discussion of possible endogeneity concerns with our probability of divorce variable and our spousal wage variables. To remind the reader, we had an

estimated probability of divorce for both our early and later periods, as well as a dummy variable indicating whether or not divorce actually occurred for our early sample. We also have for both periods a direct measure of spousal wages as well as a dummy indicating whether an individual has a higher earning spouse. As discussed earlier, we hope that our method of estimating the probability of divorce from our probit model may have helped with some of the endogeneity problems with that variable. In general though, we see that we have a problem of simultaneous causality. The theory of specialization predicts that divorce probabilities will affect wages, and it is certainly believable that wages will affect divorce probabilities. The correct way to deal with this is by finding another instrument or variable that is correlated with divorce probabilities but not with the error term in our cross sectional or longitudinal regressions. You then estimate divorce with this variable, along with other control variables. However, the difficulty of finding an instrument which satisfies these constraints can make solving this endogeneity problem difficult.

Likewise with spousal wages, if we are allowing the possibility that the wage of the spouse influences the specialization and thus wages of the husband, it is logically consistent to allow for the possibility that a husband's wages effect the specialization and wages of the wife. Thus our spousal wage variables are also endogenous. Due to the information available from the NLSY regarding one's spouse, we were not able to tackle this endogeneity problem. The NLSY simply does not have much information regarding one's spouse. In fact, spousal earning seems to be one of the few things we are told from the NLSY. Thus finding a suitable instrument to solve this endogeneity problem was not feasible.

However, we have also included regressions, as can easily be seen from previous sections, that don't include these particular endogenous variables. Furthermore, depending on how we

control for these variables, our overall wage premium can resemble the predicted premium from regressions without these variables or can be vastly different. The cases where including these variables has a large effect on the wage premium should cause concern. For this reason, we attempted to shy away from drawing major conclusions from the results of these regressions. We did however still think it was of value to note when the results were consistent with specialization or selection.

If there is one good thing, it is that the central conclusions from our previous sections are all supported by regressions not including these endogenous variables, and so we are hopeful that if a better job of solving these endogeneity problems was done, that our main conclusions would still be the same.

Lastly, we will discuss possible endogeneity problems that some of our other control variables may have. Again, whenever one tries to explain a complicated real-world process by running a regression with an imperfect or incomplete selection of control variables, there will always be questions of endogeneity problems such as omitted variable bias and simultaneous causality. As far as our paper goes, there a few control variables that could possibly be endogenous. We will discuss some of the more obvious cases below.

*Weeks worked & hours worked* – Most likely some people choose to take jobs that pay less but are enjoyable, with the intention of working more hours/weeks. This may cause our weeks worked and hours worked variables to be endogenous, as in certain cases low wages can be an incentive for working longer (or at least different) hours.

*Union Membership and Industry of Employment* – It is very possible that people with low levels of productivity or with little potential to become productive are generally attracted or

forced into certain professions. Thus it is not clear which way causality runs with these variables. We can think of one's occupation determining at least partially one's productivity, however it is also possible to think the other way around: mainly that potential wages are partially what cause the industry of employment. Many times we would expect an individual to choose an occupation specifically because of their future wage rates in that profession. This means that union membership and occupation are likely not exogenous variables.

*Children* - We would think that omitted variables could drive both the number of children one has and the wage that individual earns in the market. For example, an individual who gladly takes on responsibility may decide to have more children, and such a trait may also cause him to have a high wage. Generally speaking however, some of these variables may be constant over time though, and thus controlled for in our longitudinal regressions. However, simultaneous causality may also be present to some degree. Higher wages make it easier to raise children, and having children likely affects patterns of investment in human capital (albeit it an uncertain way. You may spend less time in the marketplace if you have a child at home. Alternatively, you may invest more time in gaining human capital as children are expensive to raise and you may need the extra market related human capital).

Other stories could likely be told to raise endogeneity questions among even more our control variables. However, there is some reason to think that this is not a very large problem. Specifically we can look at regression results under different selections of control variables. If the wage premium seems to stay fairly constant while we keep adding more and more control variables, then we would expect that while we may have some endogenous control variables, these problems are not influencing the conclusions we draw from the data in a negative way.

Indeed, though these regression results are not shown in the body of this paper (as they have little to do with the theory of specialization or the selection hypothesis) we do find that when we run regressions under different selections of control variables, that our results do not change a great deal. In any case, they do not change so much that our interpretation of the results in section VII would be changed.

In conclusion, we have discussed three areas for concern with our results. We found that we had reason for controlling for age to the extent that we did, even though ideally we would be able to better control for age while keeping our results easy to interpret. Also, while divorce probabilities and spousal income are both endogenous variables, we have dealt with them as best we could given the limits of the NLSY79 data set. In any case, our main conclusions are supported by regressions not including divorce probabilities or spousal income, so hopefully the endogeneity of these variables is not affecting our overall interpretation of the data. Lastly, while some of our other control variables may be endogenous as well, by running regressions under many different selections of control variables, and seeing that our overall interpretation of the data doesn't change, we can be hopeful that these specific endogeneity problems are not too much to worry about.

Now we are ready to conclude this paper with a reminder of what we were aiming to achieve when we set out to conduct this study, and what results we ended up finding. We will also devote some discussion to areas for future research on this topic.

## Section IX – Conclusion: A Summary of Findings and Suggestions for Future Research

We began this paper by noting that throughout the literature on labor and family economics it has been found that married men earn more than non-married men even when one controls for a host of personal characteristics. We also noted that estimates for how wide this earnings gap was seemed to change over time. It was very large in the 50's, 60's, and 70's but by the early 90's it had shrunk substantially (Korenman & Neumark(91), Gray(97)). Since then, there have been few papers written on this difference in earnings, called the "marital wage premium". Our hope was to add to the existing literature on this premium in a few key ways.

One was to bring updated results to the table. We were able to use data as recent as 2006 to estimate the wage premium.

Another hope was that we could more accurately test for the reasons behind the wage premium than previous studies. This is because previous studies failed to control for divorce probabilities and spousal income when estimating the premium: two key variables whose effects on the wage premium can be predicted in fairly precise ways by the current theories used to explain the premium. This allows us to get a more accurate picture of what is causing the wage premium.

Thirdly, we wanted to see whether or not the marital wage premium continued to decrease since the early 90's.

We answered these questions by using data from the NLSY79. We restricted our analysis to non-Black and non-Hispanic males to keep our results in line with previous studies that have mainly focused their analysis on white males. We ultimately decided to analyze the premium

experienced by a group of men from 1990-1992, and then follow that up with an analysis of the premium for that same group of men from 2002-2006. We analyzed two separate periods so that we could get a better idea of how the marital wage premium has changed over time. Ideally we would have been able to analyze the wage premium for similarly aged men over this same period; however this was not possible given the data we had. We used a probit model to estimate a probability of divorce term, and we used two specifications to estimate the wage premium: a cross sectional regression that doesn't control for fixed effects, and a longitudinal specification that does.

We found that the wage premium in 1990-1992 was extremely small after one controls for fixed effects. This suggested that the majority of the wage premium at this time was due to selection, a finding supported by other studies in the field (Gray 97). We found however, that what premium did exist after controlling for fixed effects was well explained by the theory of specialization, which is discussed in detail in section II of our paper. Our conclusion was that selection is responsible for the majority of the premium during the years 1990-1992, but that specialization also plays a small role.

This can be contrasted with our results from the years 2002-2006, which find a wage premium around 7% to 8% even after controlling for fixed effects, which is fairly substantial. In addition, selection still seems to explain the majority of the premium found, with almost no evidence supporting the idea that specialization is having an important effect on wages in this time period. We conclude from this that the majority of the later premium is due to selection with a bit of it not adequately explained by current theory.

Our ultimate conclusion was that we see an increasing wage premium since the early 1990's that can increasingly be attributed to selection, with other explanations or theories needed to explain the premium that remains after we control for selection. This result was surprising as we had thought it likely that the wage premium would continue to decrease if not totally disappear. Though we controlled for age to some extent, this result still could have been a result of the different age distributions in our early and later time period. We also don't have any obvious explanation for the apparent increase in the importance of selection.

There were also some endogeneity concerns with our variables, and it is possible, as just stated, that by failing to adequately control for the vast age difference between our early and later sample of men that our estimate of the change in the wage premium is misguided. However, from our discussion in section VIII we are somewhat optimistic that our central findings would not change if most of these endogeneity problems were removed.

Lastly, there is still much potential for future research in this area. For one, it would be worthwhile to get data which allows for a comparison of the wage premium over time for samples of similarly aged men. Fixing the endogeneity problems associated with divorce probabilities and spousal wages would also yield results beneficial to our understanding of the wage premium. It would also be worthwhile to investigate how the marital wage premium has changed for ethnic groups other than white males, or for society at large for that matter. Furthermore, it would be interesting to see a similar study done for females. Is marriage still correlated with lower wages for women? Why or why not? And how have the reasons behind the "female" marital wage premium changed over time. It would also be interesting to see how the results of a study such as this would compare to results for similar studies on men. For instance, do we also see a decline in the importance of specialization and an increase in the importance of

selection? One last project that would contribute greatly to the existing literature would be to attempt to test the theory of favoritism, which is discussed shortly in section II of this paper. Currently very few have tackled this issue (Korenman & Neumark, 91), and convincing results regarding the importance of this theory have been hard to come by.

In conclusion, while this paper has made a contribution to the existing literature on the marital wage premium by using recent data and attempting to test for specialization more extensively, there are still many questions regarding this topic that hopefully will be answered by future researchers.

Works Cited


Becker, Charles S. A Treatise on the Family. Cambridge: Harvard University Press, 1981.

Blackburn, M., & Korenman, S. (1994). The Declining Marital-Status Earnings Differential. *Journal of Population Economics, 7*(3), 247-270.

Chun, H. L., I. (2001). Why Do Married Men Earn More: Productivity or Marriage Selection? *Economic Inquiry, 39*(2001), 307 - 319.

Ginther, D. K. Z., Madeline. (2001). Is the male marriage premium due to selection? The effect of shotgun weddings on the return to marriage. *Journal of Population Economics, 14*(2), 313-328.

Gray, J. S. (1997). The Fall in Men's Return to Marriage: Declining Productivity Effects or Changing Selection? *The Journal of Human Resources, 32*(3), 481-504.

Hellerstein, J. K., Neumark, D., & Troske, K. R. (1999). Wages, Productivity, and Worker Characteristics: Evidence from Plant-Level Production Functions and Wage Equations. *Journal of Labor Economics, 17*(3), 409-446.

Hersh, J. S., Leslie S. (2000). Household specialization and the male marriage wage premium. *Industrial and Labor Relations Review, 54*(1), 78-94.

Hewitt, B. W., M. Baxter, J. (2002). Marriage and Money: The Impact of Marriage on Men's and Women's Earnings

Hill, M. S. (1979). The Wage Effects of Marital Status and Children. *The Journal of Human Resources, 14*(4), 579-594.

Korenman, S. (1994). The Declining Marital-Status Earnings Differential. *Journal of Population Economics, 7*(3), 247 - 270.

Loh, E. S. (1996). Productivity Differences and the Marriage Wage Premium for White Males. *The Journal of Human Resources, 31*(3), 566-589.

Loughran, D. S. (2002). The Effect of Male Wage Inequality on Female Age at First Marriage. *The Review of Economics and Statistics, 84*(2), 237-250.

National Center for Health Statistics, C. (2009). Provisional number of divorces and annulments and rate: United States, 2000-07: CDC.

Neumark, S. K. a. D. (1991). Does Marriage Really Make Men More Productive. *The Journal of Human Resources, 26*(2), 282-307.

Wood, R. G. (1995). Marriage Rates and Marriageable Men: A Test of the Wilson Hypothesis. *The Journal of Human Resources, 30*(1), 163-193.