

Marriage Dynamics, Earnings Dynamics, and Lifetime Family Income¹

Joseph G. Altonji* and Ivan Vidangos**

February 15, 2014

¹*Yale University and NBER. **Federal Reserve Board. Our research has been supported by the Cowles Foundation and the Economic Growth Center, Yale University (Altonji). This work was supported in part by the facilities and staff of the Yale University Faculty of Arts and Sciences High Performance Computing Center, and by the National Science Foundation under grant #CNS 08-21132 that partially funded acquisition of the facilities. The views expressed in the paper are our own and not necessarily those of the Federal Reserve Board, Yale University, NBER, or other members of their staffs. We are responsible for the remaining shortcomings of the paper. **This is a very preliminary and incomplete working draft. The empirical results are not for quotation.**

Abstract

We examine what determines the family income that individuals experience over their adult lives. To this end, we estimate a dynamic model of earnings, marriage, and divorce. The model also includes fertility. We explore the determinants of the earnings over a career of single and married men and women using a model of wages, employment, work hours, and earnings. We use the model to address a number of important questions in labor and family economics, including the effects of education and unobserved permanent characteristics on marital status and on spouse characteristics conditional on marriage. We also explore the dynamic response of wage rates, hours, earnings, marriage, and spouse characteristics and family income to various shocks and measure the relative contributions of the shocks to the variance of family income in a given year and over a lifetime.

1 Introduction

An individual's own earnings capacity has an obvious direct effect on the resources available to him/her. But it has also long been recognized that marriage plays a central role in determining the economic well being of an individual. Resources are shared among family members, so the income of other members of his/her family matter. Consequently, whether one marries and whom one marries, are key sources of variation in income of an individual over time, especially when one takes account of scale economies in household consumption. Indeed, institutions such as child support and alimony payments arise in part as a mechanism to protect against the risks to an individual of losing a valuable economic partner.

There is a close connection between an individual's earnings capacity and the income they are likely to be able to obtain access to through marriage. For example, in the data from the Panel Study of Income Dynamics (PSID) used in this study, the regression coefficient between years of education of married husbands and wives is about 0.63. The corresponding coefficient relating the husband's log hourly wage rate to the wife's hourly wage rate in cases in which both partners work is 0.401 when education is excluded and 0.295 when the education is held constant.¹ Marriage rates and marriage patterns are an important determinant of the distribution of household incomes in the cross section. Indeed, a number of papers, such as Kremer (1997), Fernandez and Rogerson (2001), and Fernandez, Guner, and Knowles (2005), study the connection between trends in patterns of assortative mating and trends in inequality. Others investigate the effect of education and earnings on the probability of marriage and on the human capital of one's spouse. The evidence of positive assortative mating indicates that permanent advantages such as education or innate ability influence an individual's economic circumstances through marriage prospects and not only through own earnings. Luck in finding a good job match early in a career may not only raise future earnings, but may enable an individual to attract a more highly skilled spouse. By the same token, a layoff may also lead to reduced income through the marriage channel in addition to its short- and long-term effects on work hours and wage rates. Furthermore, divorce has become quite common over the past four decades in many developed countries, as has remarriage. Consequently, it's not only the degree of assortative matching in marriage that is important, but also the persistence of marriage, particularly to the extent that divorce is related to the degree of mismatch in the labor market skills of a couple.

In this paper, we provide a dynamic model of earnings, marriage, and family income, and use it to address three sets of questions. The first concerns the marriage matching process. Our marriage

¹The education regression controls for year dummies and a cubic in age of both partners. The wage regressions control for year dummies and a cubic in the potential experience of both partners. Dropping these controls makes little difference.

model includes equations characterizing marital search. These specify the effects of individual characteristics on the distribution of characteristics of the potential partners the individual meets. It also includes a marriage equation relating marriage to the match between the individual and the potential partner drawn. Labor market shocks influence the type of person a single person meets and whether a marriage forms. The divorce equation relates the divorce probability to both fixed and time-varying characteristics of the marriage partners. We use the model to study how unobserved heterogeneity, education, and wage and employment shocks influence the joint distribution of marriage and marriage partner characteristics over a lifetime. We provide a richer description of marital sorting than exists in the previous literature.

The second question concerns the dynamics of earnings, family formation, and family income. In our framework, “family”, is the family an individual is in at a point in time. We build a model of individual earnings dynamics that depends on gender and marital status. We also estimate a simple process for unearned income. Combining the model of individual earnings and unearned income with the model of marriage, we can estimate how employment shocks and wage shocks influence an individual’s family income through effects on own earnings, spouse’s earnings, and marriage.

The third question is what are the sources of inequality in lifetime family income. We decompose the variance in the sum of family income per adult equivalent from age 25 to 62 into the contributions of education, fixed unobserved heterogeneity, persistent wage shocks and employment shocks, persistent wage and employment shocks affecting one’s spouse, random variation in marriage partner characteristics, and divorce shocks.

Our paper builds on several literatures. Browning, Chiappori, and Weiss (forthcoming) survey an extensive literature on marriage and divorce in an environment when search costs are relatively low, including the seminal contributions of Becker (1973, 1974, 1981) subsequent papers such as Becker, Landes and Michael (1977), Weiss and Willis (1993), Choo and Siow (2006), Chiappori and Oreficce (2008) and Chiappori, Iyigun and Weiss (2009). This literature explores the implications of comparative advantage within the family and of competition in the marriage market for who gets married, and who marries whom. Mortensen (1988), Burdett and Coles (1997, 1999), Shimer and Smith (2000), and Jacquemet and Robin (2012) are part of a literature that consider assortative matching and marriage when search costs are substantial. Wong (2003) estimates a structural model of marital sorting. The theoretical framework underlying her paper draws on the work by Burdett and Coles, which also provides motivation for our equations for marital sorting and marriage formation. The connection between the theory and the econometric specification is much looser in our model. We also allow the characteristics of the people one meets to depend on one’s own characteristics, while she assumes random meetings but implicitly allows the degree of friction in the market to differ by group.

Our work is also related to a very large literature on earnings, and earnings dynamics. Some of this work focuses on labor supply. A large number of papers estimates univariate processes for earnings and/or family income, often with a focus on implications for inequality.² A smaller set of papers investigates multivariate processes for earnings, with equations for employment, hours, and wages, and in some cases attention to job mobility.³ Our approach is closest in spirit to that of Altonji, Smith, and Vidangos (2013; hereafter, ASV). They formulate and estimate a rich model of the earnings process of male heads of household. They use the model to examine how earnings, hours, and wages respond to various shocks. They also provide decompositions of the variance of lifetime male earnings. ASV’s earnings model includes equations for employment transitions, job mobility, a wage process that depends upon new job offers, unemployment shocks, and persistent changes in general productivity, and an hours equation. The equations are not structural, but may be viewed as approximations to the decision rules relating choices to the state variables that determine them. Our approach is similar in spirit, in that we directly parameterize the key equations in our model rather than derive them as solutions to dynamic optimization problems. In comparison to their paper, we have simplified the earnings process in a number of important ways. In particular, we do not include equations for job mobility, or separate out job-specific wage and hours components. However, we consider the earnings of women as well as men, have changed the focus from an individual’s earnings to family income, and have expanded the analysis to incorporate marriage and fertility. These changes dramatically complicate the analysis. In what we do, modeling how an individual’s own characteristics influence transitions into and out of marriage is not enough. One must also model the distribution of all of the spouse characteristics that determine her earnings, including education, age, an unobserved permanent productivity component, and children. Additionally, women’s labor market behavior is strongly influenced by marriage and by children, and so it is necessary to incorporate children into the equations of the earnings model. And furthermore, children directly affect resources available to their parents, and so in considering the distribution of family income it’s important to keep track of the demand for resources posed by other family members.

We estimate our model using panel data from the PSID on labor market variables, education, marital status, and fertility. We also use a corresponding set of information on spouses which is available during the years that they are married or cohabiting with the PSID sample member. We estimate the model by indirect inference—a simulation-based methodology that has been used in

²Recent examples with references of the literature include Haider (2001), Baker and Solon (2003), Guvenen (2009), Meghir and Pistaferri (2004), DeBacker et al (2013), and Moffitt and Gottschalk (2011).

³Multivariate models of earnings dynamics include Abowd and Card (1987, 1989); Altonji, Martins, and Siow (2002); Low, Meghir and Pistaferri (2010); and Altonji, Smith and Vidangos (2013). Several recent papers do not focus on earnings but provide structural models of wage rates, job mobility, and employment dynamics. These including Barlevy (2008), Buchinsky et al (2010), and Bagger et al (2013) among others.

a number of recent papers. The fact that the model is nonlinear and involves unobservables and dynamics, and the fact that we do not see the initial conditions for many sample members and that the data are highly unbalanced makes standard method of moment or MLE approaches infeasible. As in ASV, we use a smoothing strategy that allows us to handle the presence of discrete variables in our model such as employment status, marital status, and fertility.

The remainder of the paper is organized as follows. In Section 2 we present the model. Section 3 discusses the data. Section 4 describes our estimation methodology. Section 5 provides a brief description of our preliminary estimation results, and of the analysis that we plan to conduct using the estimated model. Section 6 concludes. **As we highlighted in the Title Page note, this is a preliminary and incomplete working draft.**

2 A Model of Earnings, Marriage, and Family Income

2.1 Brief Overview

Our model consists of four parts. The first is a model of earnings of single and married men and women. The model includes equations for employment transitions, wage rates, work hours, and earnings. These processes depend on both marital status and the presence of children. The second part is a model of marriage and divorce. It is motivated by a marital search framework where search is directed. In the model single individuals meet, each period, a potential spouse with a set of characteristics that matter for marriage formation, earnings, and fertility. The distribution of the characteristics of the potential spouse, which include age, education, unobserved heterogeneity, and a persistent wage component, depend on the individual's characteristics. The probability that a marriage forms depends on the *levels* of the characteristics of both potential partners. It also depends on the *match* between them, as suggested by marriage market competition. Once formed, the probability that a marriage continues depends on the fixed characteristics of both spouses, as well as the processes that govern their earnings and fertility.

The third and most simplistic part of the model concerns fertility. Fertility depends on age, education, gender, marital status, the presence of other children, and unobserved heterogeneity. Children affect employment, hours, wages, and the divorce probability, but the model is recursive in that labor market variables do not influence fertility.

The fourth part of the model consists of equations for unearned income and an accounting identity relating family income to earned income of the individual, the earnings of his or her spouse for those who are married, and unearned household income.

We restrict the sample to individuals who are between the ages of 25 and 62. We do this because labor market information is somewhat limited for persons who are not heads of household

or wives, and in the empirical work we limit the sample to PSID heads or wives. (In the PSID the male marriage partner is considered the head in almost all cases.) Since single adults are not considered heads of household until they leave their parents for the first time, and many do not do so until their mid-to-late 20s, we focus on ages from 25 on. Living with one's parents is very common for ages younger than 25.

2.2 Model of Earnings, Unearned Income, and Family Income

The earnings model consists of equations for initial employment status and wage rates at age 25, equations for employment transitions for those employed in the previous period and those not employed in the previous period, equations governing the evolution of wage rates, an equation for work hours given employment status, the hourly wage and other variables; and an equation relating annual earnings to the wage rate, hours, and an addition stochastic component. In contrast to ASV, we abstract from job mobility and the presence of job-specific wage and hours components. These are quite important empirically, and no conceptual difficulties arise in working with a richer earnings model that incorporates these elements. However, given the complexity of dealing with marriage and children and the extra parameters needed to accomodate gender in the earnings model, we chose to simplify the model along that dimension. Additionally, as will become clear below, we have to model the distribution of all spouse characteristics that enter the earnings process. Consequently, one would have to model the distribution of the job-specific wage and hours components and job seniority for potential spouses, if one were to introduce job mobility in an interesting way.

A word about notation first. The subscript i indexes the individual. In our empirical application, i will refer to the PSID sample members. The variable t_i , which we sometimes suppress, refers to the age of the individual. We sometimes write out age_{it} even though there is some redundancy. The variable PE_{it} is years of potential labor market experience of i for a particular observation. It is a function of age (t_i) and years of education $educ_i$.⁴ The γ parameters refer to intercepts and to slope coefficients. For each intercept and slope parameter the superscripts identify the dependent variable. The subscripts of slope parameters identify the explanatory variable. We use δ to denote coefficients on the fixed person-specific unobserved heterogeneity components μ_i and η_i . The superscripts for the δ parameters denote the dependent variable and the subscripts μ and η identify the heterogeneity component. We use ρ with appropriate subscripts to denote autoregression coefficients. The ε_{it}^k are i.i.d. $N(0, \sigma_k^2)$ random variables where k corresponds to the dependent variable affected directly by ε_{it}^k .

⁴Wages, hours, and earnings are net of economy-wide year effects which we remove using a regression procedure discussed in section 4.

2.2.1 Log Hourly Wages

The log wage rate $wage_{it}$ is determined by the following system of equations:

$$\begin{aligned}
(1) \quad wage_{it} &= E_{it} \cdot wage_{it}^{lat} \\
(2) \quad wage_{it}^{lat} &= [X_{it}\gamma_X^w + \gamma_{PE^3}^w PE_{it}^3] + \gamma_{MAR}^w MAR_{it} + \gamma_F^w F_i + \gamma_{M_F}^w MAR_{it} F_i + \\
&\quad + \gamma_{F_CH}^w F_i CH_{it} + \gamma_{F_M_CH}^w MAR_{it} F_i CH_{it} \\
&\quad + \delta_\mu^w \mu_i + \omega_{it} \\
(3) \quad \omega_{it} &= \rho_\omega \omega_{i,t-1} + \gamma_{1-E_{it}}^\omega (1 - E_{it}) + \gamma_{1-E_{it-1}}^\omega (1 - E_{i,t-1}) + \varepsilon_{it}^\omega \\
\varepsilon_{it}^\omega &\sim N(0, \sigma_\omega^2)
\end{aligned}$$

Equation (1) says that for employed individuals (i.e. $E_{it} = 1$), $wage_{it}$ equals the “latent wage” $wage_{it}^{lat}$. For an individual who is not employed, $wage_{it}^{lat}$ captures the process for wage offers. At a given point in time the individual might not have such an offer. The formulation parsimoniously captures the idea that worker skills and worker-specific demand factors evolve during a nonemployment spell. It also allows us to deal with the fact that in the data wages are only observed for jobs that are held at the survey date.

Equation (2) states that $wage_{it}^{lat}$ depends on four components. The first is $X_{it}\gamma_X^w + \gamma_{PE^3}^w PE_{it}^3$, where X_{it} is a vector of exogenous variables consisting of the race indicator $BLACK_i$, years of education $EDUC_i$, a quadratic in PE_{it} , and a constant, and PE_{it}^3 is the cube of experience. Since we do not control for tenure effects or gains from job shopping, the coefficients on the powers of PE pick up these effects, in addition to the effect of general human capital and/or an aging effect. The third term is the permanent unobserved “ability” component μ_i .

The fourth term is a stochastic component ω_{it} , which according to (3) depends on $\omega_{i,t-1}$, the current value and the first lag of nonemployment ($1 - E_{it}$), and the error component ε_{it}^ω . The dependence of ω_{it} on its past reflects persistence in the market value of the general skills of i and/or the fact that employers base wage offers on past wages. It will also pick up persistence arising from job-specific wage components that change slowly within an employer-employee match.

2.2.2 Employment-Employment Transitions (EE_t)

The dummy EE_{it} indicates whether a worker who was employed in $t - 1$ remains employed in t . It is determined by

$$\begin{aligned}
(4) \quad EE_{it} = & I[X_{i,t-1}\gamma_X^{EE} + \gamma_M^{EE}MAR_{it} + \gamma_F^{EE}F_i + \gamma_{M_F}^{EE}MAR_{it}F_i \\
& + \gamma_{\hat{w}}^{EE}wage'_{it} + \gamma_{M_F_w}^{EE}MAR_{it}F_iwage'_{it} + \gamma_{\hat{w}_s}^{EE}MAR_{it}F_iwage'_s_{it} \\
& + \gamma_{F_CH}^{EE}F_iCH_{it} + \gamma_{F_M_CH}^{EE}MAR_{it}F_iCH_{it} \\
& \delta_\mu^{EE}\mu_i + \delta_\eta^{EE}\eta_i + \varepsilon_{it}^{EE} > 0]
\end{aligned}$$

where $I(\cdot)$ is the indicator function, and $wage'_{it}$ is what the wage would be in t if the individual were to continue employment. The variable $wage'_{it}$ is the value of $wage_{it}^{lat}$ determined by (2) and (3) with $E_{it} = 1$ and $E_{i,t-1} = 1$. The vector $X_{i,t-1}$ is the same as X_{it} except that it contains $PE_{i,t-1}$ and $PE_{i,t-1}^2$ rather than PE_{it} and PE_{it}^2 . In ASV, we included employment duration in the model, but here we leave it out for simplicity. Standard labor supply models imply that employment at t should depend on the current wage opportunity, which we proxy with $wage'_{it}$. We also allow employment of married women to depend on their spouse's wage if they remain employed, $wage'_s_{it}$. EE_{it} also depends on the permanent ability component μ_i as well as a second heterogeneity component η_i (which we will refer to as “propensity to move”). The model allows children CH_{it} to affect employment transitions of women, with the size of the effect depending on marital status. We impose the restriction that the employment and hours of men do not depend on the wage of their spouse, because employment and hours regressions did not reveal a strong link.

Both permanent unobserved heterogeneity components μ_i and η_i also directly affect transitions out of employment, marriage, and work hours, but η_i is excluded from the wage model. One may think of η_i as a factor that is related to labor supply and to employment and marriage preferences but not productivity. The current empirical version of the model only includes one source of heterogeneity— μ_i .

2.2.3 Nonemployment-Employment Transition (NE_t):

Movement from nonemployment to employment is determined by

$$\begin{aligned}
(5) \quad NE_{it} = & I[X_{i,t-1}\gamma_X^{NE} + \gamma_{MAR}^{NE}MAR_{it} + \gamma_F^{NE}F_i + \gamma_{M_F}^{NE}MAR_{it}F_i \\
& + \gamma_{F_CH}^{NE}F_iCH_{it} + \gamma_{F_M_CH}^{NE}MAR_{it}F_iCH_{it} \\
& + \gamma_{w_s}^{NE}F_iwage'_s_{it} + \gamma_{\hat{w}}^{NE}wage'_{it} + \gamma_{NED}^{NE}NED_{i,t-1} \\
& + \delta_\mu^{NE}\mu_i + \delta_\eta^{NE}\eta_i + \varepsilon_{it}^{NE} > 0]
\end{aligned}$$

where $NED_{i,t-1}$ is the number of years of nonemployment at the $t - 1$ survey date, and $NED_{it} = (1 - E_{it})(NED_{i,t-1} + 1)$. So far, we have restricted the duration dependence parameter γ_{NED}^{NE} to 0 in the empirical work. The transition equation includes the heterogeneity components μ_i and η_i .

Note that E_{it} is given by

$$E_{it} = EE_{it}E_{i,t-1} + NE_{it}(1 - E_{i,t-1}).$$

2.2.4 Log Annual Hours

Log annual work hours are determined by

$$\begin{aligned}
(6) \quad hours_{it} = & \gamma_0^h + \gamma_X^h X_{it} + \gamma_{pe^3}^h PE_{it}^3 + \gamma_{MAR}^h MAR_{it} + \gamma_F^h F_i + \gamma_{M_F}^h MAR_{it} F_i \\
& + \gamma_{F_CH}^h F_i CH_{it} + \gamma_{F_M_CH}^h MAR_{it} F_i CH_{it} \\
& + \gamma_E^h E_{it} + \gamma_{F_E}^h FEM_{it} E_{it} + \gamma_{F_M_E}^h MAR_{it} FEM_{it} E_{it} \\
& + \gamma_w^h wage_{it}^{lat} + \gamma_{F_w}^h F_i wage_{it}^{lat} + \gamma_{F_M_ws}^h MAR_{it} F_i wage_{it}^{lat} \\
& + \delta_\mu^h \mu_i + \delta_\eta^h \eta_i + \varepsilon_{it}^h
\end{aligned}$$

The $hours_{it}$ equation includes X_{it} and PE_{it}^3 . It also includes η_i and μ_i . The i.i.d. error component ε_{it}^h picks up transitory variation in straight time hours worked, overtime, multiple job holding, and nonemployment conditional on employment status at the survey. It may reflect temporary shifts in worker preferences as well as hours constraints. As noted earlier, we exclude a job-match-specific component from the hours equation despite evidence in ASV and other studies that such components are important.

Hours also depend on $wage_{it}^{lat}$ and E_{it} . For many observations, $wage_{it}^{lat}$ is the actual wage. However, many individuals who are not employed at the survey date work part of the year. For these individuals $wage_{it}^{lat}$ is the wage the individual would typically receive. Because wage shocks turn out to be highly persistent and because we strongly question the standard labor supply assumption that individuals are free to adjust hours on their main job in response to short-term variation in wage rates, we regard the coefficient on the latent wage as the response to a relatively permanent wage change rather than the Frisch elasticity. We allow the wage coefficient to depend on gender and on marital status in the case of women. Married women also respond to their spouse's wage. Children affect hours for women but not for men.

When interpreting results for EE_{it} , one must keep in mind that our employment indicator refers to the survey date. As a result, we undoubtedly miss short spells of unemployment that fall between surveys. Fortunately, however, earnings depend on employment through annual work hours, and the transitory error component in the hours equation should capture the effect on hours from unemployment or departures from the labor force of varying duration. Note that we allow the relationship between $hours_{it}$ and E_{it} to depend on F_i and $F_i MAR_{it}$. A substantial fraction of married women may be out of the labor force for the entire year.

2.2.5 Log Annual Earnings

$$\begin{aligned}
(7) \quad \text{earn}_{it} &= \gamma_0^e + \gamma_w^e \text{wage}_{it}^{lat} + \gamma_h^e \text{hours}_{it} + e_{it} \\
e_{it} &= \gamma_{MAR}^e \text{MAR}_{it} + \gamma_F^e F_i + \gamma_{M_F}^e \text{MAR}_{it} F_i + \gamma_{F_M_CH}^e \text{MAR}_{it} F_i \text{CH}_{it} \\
&\quad + \rho_e e_{it-1} + \varepsilon_{it}^e
\end{aligned}$$

Log earnings earn_{it} depends on wage_{it}^{lat} and hours_{it} . The coefficients γ_w^e and γ_h^e might differ from 1 for a number of reasons, including overtime, multiple job holding, bonuses and commissions, job mobility, and the fact that for some salaried workers the wage reflects a set work schedule but annual hours worked may vary. In the empirical work, we set γ_w^e and γ_h^e to 1.

We also include a first-order autoregressive error component e_{it} to capture some of these factors, and allow the values to depend on marital status, gender, and children. Entering these variables through e_{it} rather than directly allows for dynamic effects. Note that X_{it} is excluded but influences earnings through wage_{it}^{lat} and hours_{it} .

2.2.6 Unearned Income

The model of unearned income, unearn_{it} , is not included in this draft.

2.2.7 Family Income

The level of family income is determined by the identity

$$(8) \quad Y_{it} = \exp^{\text{earn}_{it}} + \exp^{\text{earn}_{-sit}} + \exp^{\text{unearn}_{it}}$$

This assumes that other family members, such as adult children, do not contribute income. We plan to convert family income per adult equivalent.

2.3 Marriage

The marriage model draws loosely on the two-sided search-theoretic models of the marriage market considered by Collins and McNamara (1990), Burdett and Coles (1999), Wong (2003), and others. Wong (2003) estimates a version of her model using the PSID and provides estimates of how particular traits are valued in the marriage market.⁵ In our model, single individuals draw,

⁵Browning, Chiappori, and Weiss (forthcoming) survey the literature on marriage matching, marriage formation, and divorce. [Add reference to Chiappori and Siow studies that use a frictionless equilibrium model to examine marital sorting.] There is a substantial descriptive literature analyzing the effects of personal characteristics on the probability of marriage and the effects of the characteristics of couples on the probability of divorce. Examples of studies that have examined the correlation in spouse characteristics include Spuhler (1982), Keller, Thiessen, and Young (1996).

each period, a vector of potential partner characteristics from a distribution that depends on their own characteristics. The probability that a marriage occurs depends on the potential partner's characteristics as well as one's own characteristics. Once formed, the probability that the marriage continues depends on fixed characteristics of both partners and on the lagged values of stochastic components, such as the wage, employment, and children. The joint distribution of the characteristics of the couple is determined by functions that determine potential spouse characteristics, the function that determines marriage formation, and the function that determines marriage dissolution. As we will discuss below, separately identifying these functions in the absence of information on potential marriages that do not form is a challenge.

Note that the parameters of the model depend on the supply of men and women in the marriage market, and well as the distribution of preferences over the characteristics of partners and the value of being married relative to single life. They are not likely to be invariant to changes in divorce laws, tax policy, labor market discrimination, changes in age structure, and changes in preferences for children.

For the marriage transition equations it is helpful to define variables corresponding to the male and the female in the couple that person i is in. For a generic variable Z_i and the corresponding spouse value Z_{-s_i} , let Z_{-M_i} denote the values for the male and Z_{-F_i} denote the value for the female in the couple. For example, AGE_{-M_i} and AGE_{-F_i} are the ages of the male and the female. In terms of the prior notation, $Z_{-M_i} = [(1 - F_i)Z_i + F_iZ_{-s_i}]$ and $Z_{-F_i} = [(1 - F_i)Z_{-s_i} + F_iZ_i]$.

For individuals who are married at $t - 1$, the continuation of the marriage into period t is determined by

$$\begin{aligned}
(9) \quad MAR_{it} = & I[\gamma_0^{MM} + \gamma_{AGE_{-F}}^{MM} AGE_{-F_{i,t-1}} + \gamma_{AGE_{-M}}^{MM} AGE_{-M_{i,t-1}} \\
& + \gamma_{EDUC_{-F}}^{MM} EDUC_{-F_{i,t-1}} + \gamma_{EDUC_{-M}}^{MM} EDUC_{-M_{i,t-1}} \\
& + \gamma_{E_{-F}}^{MM} E_{-F_{i,t-1}} + \gamma_{E_{-M}}^{MM} E_{-M_{i,t-1}} \\
& + \gamma_{w_{-F}}^{MM} wage^{lat}_{-F_{i,t-1}} + \gamma_{w_{-M}}^{MM} wage^{lat}_{-M_{i,t-1}} \\
& + \gamma_{CH}^{MM} CH_{i,t-1} + \gamma_{MD}^{MM} MDUR_{i,t-1} \\
& + \delta_{\mu_{-M}}^{MM} \mu_{-M_{i,t-1}} + \delta_{\mu_{-F}}^{MM} \mu_{-F_{i,t-1}} + \delta_{\eta_{-M}}^{MM} \eta_{-M_{i,t-1}} + \delta_{\eta_{-F}}^{MM} \eta_{-F_{i,t-1}} \\
& + \delta_{\mathbf{v}}^{MM} \mathbf{v}_{i,t-1} + u_{it}^{MM} > 0], \text{ conditional on } MAR_{i,t-1} = 1.
\end{aligned}$$

In the above equation $I[\cdot]$ is the indicator function, $u_{it}^{MM} \sim N(0, 1)$ is i.i.d., $MDUR_{i,t-1}$ is “marriage duration”, defined as the number of years that the couple was married as of $t - 1$, and $CH_{i,t-1}$ is a summary variable that indicates the presence of young children in the marriage. In the empirical work to date, we have restricted $CH_{i,t-1}$ to be just an indicator of the presence of children of age between 0 and 1. In future versions, we will use a richer specification that allows older children

to also affect the probability of the continuation of the marriage. Factors $\mu_{-M_{i,t-1}}$ and $\mu_{-F_{i,t-1}}$ represent the permanent unobserved “ability” corresponding to the male and female, respectively. The notation indexes these factors by t . Person i ’s value of μ is time-invariant and so is the value of μ for the person i is married to. However, with divorce and remarriage, i might change partners. Factors $\eta_{-M_{i,t-1}}$ and $\eta_{-F_{i,t-1}}$ are defined similarly. The variable $\mathbf{v}_{i,t-1}$ represents unobserved match-specific heterogeneity in the value of the marriage, including love.

Next, consider individuals who are single at age $t-1$. At that age the person draws characteristics of a potential spouse. Whether they are married in period t to that person is determined by

$$\begin{aligned}
(10) \quad MAR_{it} = & I[\gamma_0^{SM} + \gamma_{AGE_F}^{SM} AGE_F_{i,t-1} + \gamma_{AGE_M}^{SM} AGE_M_{i,t-1} \\
& + \gamma_{EDUC_F}^{SM} EDUC_F_{i,t-1} + \gamma_{EDUC_M}^{SM} EDUC_M_{i,t-1} \\
& + \gamma_{E_F}^{SM} E_F_{i,t-1} + \gamma_{E_M}^{SM} E_M_{i,t-1} \\
& + \gamma_{w_F}^{SM} wage^{lat}_F_{i,t-1} + \gamma_{w_M}^{SM} wage^{lat}_M_{i,t-1} \\
& + \gamma_{CH}^{SM} CH_{i,t-1} + \gamma_{SD}^{SM} SDUR_{i,t-1} \\
& + \delta_{\mu_M}^{SM} \mu_{-M_{i,t-1}} + \delta_{\mu_F}^{SM} \mu_{-F_{i,t-1}} + \delta_{\eta_M}^{SM} \eta_{-M_{i,t-1}} + \delta_{\eta_F}^{SM} \eta_{-F_{i,t-1}} \\
& + \delta_{\mathbf{v}}^{SM} \mathbf{v}_{i,t-1} + u_{it}^{SM} > 0], \text{ conditional on } MAR_{i,t-1} = 0.
\end{aligned}$$

In future drafts, we plan to let the transition probabilities additionally depend on the absolute difference between one’s own characteristics and the characteristics of the spouse (or potential partner, in the case of the unmarried-married transition), in order to capture the effect of marriage competition and/or compatibility of spouses’ characteristics.⁶

2.4 The Distribution of Characteristics of a Potential Spouse

The attractiveness of a marriage partner involves characteristics that are relevant for earnings as well as characteristics that are not. We have to model the distributions of all characteristics that determine future values of earnings and unearned income. In a competitive marriage market, all characteristics of an individual could in principle influence the type of partners that the individual is most likely to meet and explore a relationship with. For example, we allow the distribution of the age of the potential spouse to depend on AGE_{it} and gender, even though high-income middle-aged men might be more inclined to search for younger women, holding everything else equal, than low-income middle-aged men. Our specification is as follows.

⁶In simulation experiments we had difficulty identifying the factor loadings $\delta_{\mathbf{v}}^{MM}$ and $\delta_{\mathbf{v}}^{SM}$ on $\mathbf{v}_{i,t-1}$ in the marriage transition equations. (We assumed that potential partners draw $\mathbf{v}_{i,t-1}$ from a standard normal distribution and that it is fixed within a marriage.) In principle the factor loadings would seem to be identified by the degree to which marriages with unfavorable values of the observed variables nevertheless form and endure. However, transitions into and out of marriage are the only observables that depend directly on $\mathbf{v}_{i,t-1}$. Thus far we have excluded $\mathbf{v}_{i,t-1}$ from the empirical model.

Unobserved heterogeneity:

The value of μ for the person that i meets at age t is given by

$$(11) \quad \begin{aligned} \mu_{-sit} &= \gamma_0^{\mu-s} + \delta_{\mu}^{\mu-s} \mu_i + \varepsilon_{it}^{\mu-s} \\ \varepsilon_{it}^{\mu-s} &\sim N(0, \sigma_{\mu-s}^2), \text{ where } \sigma_{\mu-s}^2 = 1 - (\delta_{\mu}^{\mu-s})^2. \end{aligned}$$

Note that we are imposing the symmetry restriction that impose that $var(\mu_{-sit}) = var(\mu_i) = 1$. Note also that μ_{-sit} is constant the particular potential partner met in period t , just as μ_i is fixed.

Education:

The education $EDUC_S$ of the potential spouse i meets in t is given by

$$(12) \quad \begin{aligned} EDUC_S_{it} &= \gamma_0^{ED-s} + \gamma_{ED}^{ED-s} EDUC_i + \gamma_F^{ED-s} F_i + \delta_{\mu}^{ED-s} \mu_i + \varepsilon_{it}^{ED-s}, \\ \text{where } \varepsilon_{it}^{ED-s} &\sim N(0, \sigma_{ED-s}^2), \text{ and } \sigma_{ED-s}^2 \text{ is a free parameter.} \end{aligned}$$

We allow for gender asymmetries with an intercept shifter, and also allow both education and μ_i to influence the distribution of spouse's education. Note that by definition μ_i is orthogonal to $EDUC_i$. Nevertheless, if it is valuable in the marriage market or influences tastes for spouse's education, it might influence the distribution of $EDUC_S_{it}$.

Age and potential experience:

In the marriage market individuals target the age of potential spouses. We use a simple specification, in which

$$(13) \quad \begin{aligned} AGE_S_{it} &= \max\{1.0, \gamma_0^{AGE-S} + \gamma_{AGE}^{AGE-S} AGE_{it} + \gamma_F^{AGE-S} F_i + \varepsilon_{it}^{AGE-S}\} \\ \text{where } \varepsilon_{it}^{AGE-S} &\sim N(0, \sigma_{AGE-S}^2), \text{ and } \sigma_{AGE-S}^2 \text{ is a free parameter.} \end{aligned}$$

In the data, the dispersion in the gap between the husband's age and wife's age in new marriages widens with age t , suggesting that the variance parameter σ_{AGE-S}^2 increases somewhat with age t . However, in our initial explorations we ran into numerical difficulties when we allowed σ_{AGE-S}^2 to depend on t .

The value of potential experience depends on the draw of AGE_S_{it} and $EDUC_S_{it}$ through the function

$$PE_S_{it} = \max\{0, AGE_S_{it} - \max\{EDUC_S_{it}, 9\} - 6\}.$$

Employment status:

The distribution of employment for potential spouses depends on gender and their potential experience according to

$$(14) \quad \begin{aligned} E_S_{it} &= I[\gamma_0^{E-S} + \gamma_{PE_S}^{E-S} PE_S_{it} + \gamma_{F_S}^{E-S} F_S_i + \varepsilon_{it}^{E-S} > 0] \\ \varepsilon_{it}^{E-S} &\sim N(0, 1). \end{aligned}$$

Stochastic wage component:

Draws of the potential spouse's stochastic wage component are given by

$$(15) \quad \begin{aligned} \omega_S_{it} &= \gamma_0^{\omega-S} + \gamma_{\omega_S}^{\omega-S} \omega_{it} + \varepsilon_{it}^{\omega-S} \\ \text{where } \varepsilon_{it}^{\omega-S} &\sim N(0, \sigma_{\omega_S}^2), \text{ and } \sigma_{\omega_S}^2 \text{ is a free parameter.} \end{aligned}$$

Again, one could imagine that the μ_S_{it} , ω_S_{it} , $EDUC_S_{it}$, and E_S_{it} , depend on all of the state variables that influence the current and future earnings of i as of age t . The dependence might also depend on gender and age. However, some parsimony is needed.

Autoregressive component of earnings:

The distribution of the component capturing variation in earnings given the wage and annual hours is given by

$$(16) \quad e_S_{it} \sim N \left(0, \sigma_{e1}^2 * (\rho_e^2)^{\max\{0, PE_S_{it-1}\}} + \sigma_e^2 * \frac{1 - (\rho_e^2)^{\max\{0, PE_S_{it-1}\}}}{1 - (\rho_e^2)} \right),$$

where σ_e^2 is the variance of the earnings innovation and σ_{e1}^2 is the initial earnings innovation variance. The formula implied the characteristics of i influence e_S_{it} only through PE_S_{it} . Note that we usually impose the restriction that $\sigma_{e1}^2 = \sigma_e^2$.

2.5 Fertility

As noted in the introduction, the most simplistic part of the model concerns fertility. Fertility depends on marital status, gender, age, education, the presence of other children, and unobserved heterogeneity.

$$(17) \quad \begin{aligned} BirthM_{it} &= I[\gamma_0^{BM} + \gamma_{AGE_F}^{BM} AGE_F_{i,t-1} + \gamma_{AGE_M}^{BM} AGE_M_{i,t-1} \\ &+ \gamma_{EDUC_F}^{BM} EDUC_F_{i,t-1} + \gamma_{EDUC_M}^{BM} EDUC_M_{i,t-1} \\ &+ \gamma_{CH}^{BM} CH_{i,t-1} + \delta_{\mu_M}^{BM} \mu_M_{i,t-1} + \delta_{\mu_F}^{BM} \mu_F_{i,t-1} \\ &+ u_{it}^{BM} > 0], \text{ conditional on } MAR_{i,t-1} = 1, \end{aligned}$$

$$\begin{aligned}
(18) \quad BirthS_{it} = & I[\gamma_0^{BS} + \gamma_{AGE_F}^{BS} AGE_F_{i,t-1} + \gamma_{AGE_M}^{BS} AGE_M_{i,t-1} \\
& + \gamma_{EDUC_F}^{BS} EDUC_F_{i,t-1} + \gamma_{EDUC_M}^{BS} EDUC_M_{i,t-1} \\
& + \gamma_{CH}^{BS} CH_{i,t-1} + \delta_{\mu_M}^{BS} \mu_M_{i,t-1} + \delta_{\mu_F}^{BS} \mu_F_{i,t-1} \\
& + u_{it}^{BS} > 0], \text{ conditional on } MAR_{i,t-1} = 0,
\end{aligned}$$

$$Birth_{it} = MAR_{i,t-1} BirthM_{it} + (1 - MAR_{i,t-1}) BirthS_{it}$$

2.6 Initial Conditions

The model is estimated using data on PSID sample members between the ages of 25 and 62 for reasons mentioned above and discussed in more detail in the Data section. We specify the joint distribution of $E_{i,25}$, $MAR_{i,25}$, and $N_CH_{i,25}$ to depend on $EDUC_i$, FEM_i , and μ_i :

$$(19) \quad F(E_{i,25}, MAR_{i,25}, N_CH_{i,25} | EDUC_i, FEM_i, \mu_i).$$

The distribution of the ages of children is based on the empirical distribution conditional on the value of $N_CH_{i,25}$.

The initial conditions for the wage, and the autoregressive component of the wage and earnings are given by:

$$\begin{aligned}
(20) \quad Wage : wage_{i1}^{lat} = & \gamma_X^w X_{i1} + \gamma_{MAR}^E MAR_{it} + \gamma_F^E F_{it} + \gamma_{M_F}^E MAR_{it} F_{it} \\
& + \gamma_{F_CH}^w F_{it} CH_{it} + \gamma_{F_M_CH}^w MAR_{it} F_{it} CH_{it} + \omega_{i1} + \delta_\mu^w \mu_i \\
& Autoregressive : \omega_{i1} \sim N(0, \sigma_{\omega_1}^2) \\
& Earnings Error : e_{i1} \sim N(0, \sigma_e^2)
\end{aligned}$$

Since in the model ω_{it} also influences the marriage probability, it would be desirable to also jointly draw for $\omega_{i,25}$, but instead we assume that is an independent normal draw. Also, it would be desirable to allow the variance of ω_1 to depend on education and gender, but we have not done so thus far.⁷ In the empirical work so far, we exclude the unobservable μ_i from (19).

⁷For men, in a somewhat richer model of wages, ASV find that the variance of ω_1 is higher for the better educated.

2.7 Measurement Error and Observed Wages, Hours, Earnings and Unearned Income

The observed (measured) variables are:

$$(21) \quad wage_{it}^* = E_{it}(wage_{it}^{lat} + m_{it}^w)$$

$$(22) \quad hours_{it}^* = hours_{it} + m_{it}^h$$

$$(23) \quad earn_{it}^* = earn_{it} + m_{it}^e$$

$$(24) \quad unearn_{it}^* = unearn_{it} + m_{it}^{ue}$$

$$(25) \quad wage_s_{it}^* = E_s_{it} \cdot (wage_s_{it}^{lat} + m_s_{it}^w)$$

$$(26) \quad hours_s_{it}^* = hours_s_{it} + m_s_{it}^h$$

$$(27) \quad earn_s_{it}^* = earn_s_{it} + m_s_{it}^e$$

$$(28) \quad \log(Y_{it}^*) = \log(\exp^{earn_{it}^*} + \exp^{earn_s_{it}^*} + \exp^{unearn_{it}})$$

The measurement errors m_{it}^w , m_{it}^h , m_{it}^e , m_{it}^{ue} are $N(0, \sigma_{m\tau}^2)$, $\tau = w, h, e, ue$. We assume that the measurement errors in the $wage_s_{it}^*$, $hours_s_{it}^*$, and $earn_s_{it}^*$ have the same variances as the measurement errors in $wage_{it}^*$, $hours_{it}^*$, and $earn_{it}^*$. The measurement errors are assumed to be i.i.d. across i and t , mutually independent, and independent from all other error components in the model. In practice, one family member typically serves as the respondent for both the husband and wife, so some correlation among the measurement errors is likely. From the accounting identity, the measurement error in $\log(Y_{it}^*)$ is a function of the measurement errors in its components.

In our empirical work, we set the values of the variance of the various measurement error components to values suggested by various studies of measurement error in the PSID and other panel data sets, as well as by patterns in the data. See ASV.

3 Data

3.1 Panel Study of Income Dynamics

We use the 1975–1997 waves of the PSID to assemble data that refer to the calendar years 1975–1996.⁸ We use the stratified random sample (SRC). We include sample members with ages between 25 and 62, as well as their spouses (who are nonsample members). The survey covers spouses only in years in which they are married to a sample member. We restrict the age of the sample member to 25–62 because many sample members younger than 25 are neither heads nor wives, and some

⁸We start with year 1975 because the wage variables we use are not available for wives or for salaried heads of household in prior years. Extending the sample forward is complicated by the fact that the survey moved to a biennial interview schedule after 1997, but it could be done.

key variables are not collected for non-head singles. The vast majority of these individuals are children or stepchildren of the head or wife.⁹ We exclude blacks, who are underrepresented in the SRC sample. We also exclude observations where potential experience of the sample member or his spouse is greater than 40, as our model of the labor market does not cover retirement. Observations for a given person-year are used if the person had valid data on education (EDUC). Our sample includes self-employed individuals.

After imposing the restrictions above, we end up with a sample of 59,101 person-year observations on 4,945 PSID sample members. Table I shows the number of observations, the mean and the standard deviation for the key variables in our sample, for both sample members and, if married, their spouse.

3.2 Key Variables

Education is years of education. It is the average of the reports of education for heads or wives who are older than 18 when multiple reports are available. Potential experience PE_{it} is $age_{it} - \max(EDUC_i, 9) - 6$. The wage measure is the reported hourly wage rate at the time of the survey. It is only available for persons who are employed or on temporary layoff.¹⁰ The hours measure is annual hours worked in all jobs. Earnings is annual earnings in all jobs. Earning and wage rates are in year 2000 dollars. Employment status is 1 if employed or on temporary layoff and 0 otherwise. A value of 0 refers to nonemployment rather than only unemployment. The construction of marital status and marital duration are based on marital status questions that refer to the survey data as well as the PSID marital history file. The indicators for whether the individual has a child aged 0 to 1, 1 to 2, 2 to 3, etc., up to age 17, were constructed from the childbirth and adoption file. We will discuss them in a future draft.

⁹After children set up their own household, they are classified as heads or wives even if they move back in with their parents. An alternative is to start at an earlier age and treat “single, have not left home” as a state variable, and account for the fact that some labor market are not available for non-head single adults. Starting at an earlier age and restricting the sample to heads and wives will lead to bias because the marriage rate is overstated in this sample. Our results apply to individuals who are heads or married at age 25.

¹⁰This measure is the log of the reported hourly wage at the survey date for persons paid by the hour and is based on the salary per week, per month, or per year reported by salary workers. For household heads, it is unavailable prior to 1970 and is limited to hourly workers prior to 1976. We account for the fact that it is capped at \$9.98 per hour prior to 1978 by replacing capped values for the years 1975-1977 with predicted values constructed by Altonji and Williams (2005). They are based on a regression of the log of the reported wage on a constant and the log of annual earnings divided by annual hours using the sample of individuals in 1978 for whom the reported wage exceeds \$9.98. For married women the wage is available in the 1976 survey and from 1979 on, but is missing in 1977 and 1978. It is also missing for the self employed in all years. In these cases we set the wage to the log of earnings divided by hours. This introduces some correlation between the measurement error in $wage_{it}^*$ and the measurement errors in $hour_{it}^*$ and $earn_{it}^*$, which we have ignored so far.

3.3 Outliers

We censor reported hours at 4,000, add 200 to reported hours (and 1,000 to reported earnings) before taking logs to reduce the impact of very low values of hours (and earnings) on the variation in the logarithm, and censor observed earnings and observed wage rates (in levels, not logs) to increase by no more than 500% and decrease to no less than 20% of their lagged values. We also censor wages to be no less than \$3.50 in [year 2000] dollars.

For the equations in our auxiliary model that use age of the spouse as a left-hand-side variable, we censor the age of the spouse based on quantile regressions, using predictions from the 5th quantile regression and the 95th quantile regression to set bounds. That is, we do not try to match the tails of the conditional distribution of spousal age in the equations that determine spousal age as a function of the sample member characteristics.

As we have already noted, an advantage of indirect inference is that by incorporating the sample selection process into the simulation, one can handle unbalanced data. We assume that observations are missing at random, although there is reason to believe that the heterogeneity components and shocks influence attrition from the sample.

4 Estimation

[This preliminary draft of this section draws very heavily on ASV.] We estimate the model presented above using indirect inference (I-I). Our estimation approach builds on the approach in ASV. However, because of the complexity of the model, we do not use a simple seemingly unrelated regressions (SUR) formulation as the main component of the auxiliary model, as in ASV. Instead, our auxiliary model consists of a series of separate SUR systems, each tailoring a particular part of the structural model. This section provides a brief overview of our estimation procedure. The next section defines our auxiliary model.

For clarity, we refer to the model presented in previous sections as the “structural” model, even though the model does not express the parameters of the decision rules for employment participation, marriage, or fertility in terms of preference parameters and parameters representing specific constraints in the economic environment faced by the agents. We denote the k “structural” parameters by β . I-I involves the use of an “auxiliary” statistical model (with p parameters θ satisfying $p \geq k$) that captures properties of the data. The method involves simulating data from the structural model (given a hypothesized value of β) and choosing the estimator $\hat{\beta}$ of β to make the simulated data match the actual data as closely as possible according to some criterion that involves θ .

Suppose that the observed data consist of a set of observations on N individuals in each of

T periods: $\{y_{it}, x_{it}\}$, $i = 1, \dots, N$, $t = 1, \dots, T$, where y_{it} is endogenous to the model and x_{it} is exogenous. The auxiliary model parameters θ can be estimated using the observed data as the solution to:

$$\hat{\theta} = \arg \max_{\theta} \mathcal{L}(y; x, \theta),$$

where $\mathcal{L}(y; x, \theta)$ is the likelihood function associated with the auxiliary model, $y \equiv \{y_{it}\}$ and $x \equiv \{x_{it}\}$.

Given x and assumed values of β , we use the structural model to generate M statistically independent simulated data sets $\{\tilde{y}_{it}^m(\beta)\}$, $m = 1, \dots, M$. Each of the M simulated data sets has N individuals and is constructed using the same observations on the exogenous variables, x . For each of the M simulated data sets, we compute $\tilde{\theta}_m(\beta)$ as

$$\tilde{\theta}_m(\beta) = \arg \max_{\theta} \mathcal{L}(\tilde{y}_m(\beta); x, \theta),$$

where the likelihood function associated with the auxiliary model is evaluated using the m th simulated data set $\tilde{y}_m(\beta) \equiv \{\tilde{y}_{it}^m(\beta)\}$ rather than the real data. Denote the average of the estimated parameter vectors by $\tilde{\theta}(\beta) \equiv M^{-1} \sum_{m=1}^M \tilde{\theta}_m(\beta)$.

I-I generates an estimate $\hat{\beta}$ of the structural parameters by choosing β to minimize the distance between $\hat{\theta}$ and $\tilde{\theta}(\beta)$ according to some metric. We choose $\hat{\beta}$ to minimize the difference between the constrained and unconstrained values of a pseudo likelihood function of the auxiliary model evaluated using the observed data. In particular, we calculate

$$\hat{\beta} = \arg \min_{\beta} [\mathcal{L}(y; x, \hat{\theta}) - \mathcal{L}(y; x, \tilde{\theta}(\beta))].$$

Gourieroux, Monfort, and Renault (1993) have shown that $\hat{\beta}$ is a consistent and asymptotically normal estimate of the true parameter vector β_0 .

Accommodating missing data in I-I is straightforward: after generating a complete set of simulated data, one simply omits observations in the same way in which they are omitted in the observed data. We assume that the pattern of which years data are available on sample members is exogenous. In some cases, it is convenient to estimate auxiliary models in which missing observations are replaced with some arbitrary value such as 0 or the sample mean. In such circumstances, the same principle applies: use the same arbitrary values in both the simulated and observed data sets. The fact that we observe people in the data for the first time at various ages would pose extremely serious “initial conditions” problems if we were using standard panel data methods, but is handled naturally by I-I because the missing early observations will affect the probability limits of $\tilde{\theta}(\hat{\beta})$ and $\hat{\theta}$ in the same way.¹¹ However, thus far we have not dealt with the fact that availability of wage

¹¹Heckman (1981) and Wooldridge (2005) suggest dealing with initial conditions by using a flexible form for the distribution of the first observation for each i and its relationship to error distributions in the outcome equations for

data in 1975, 1977, and 1978 depends on marital status in the case of females. Doing so will not be difficult.

The presence of discrete random variables complicates the search for $\hat{\beta}$ because the objective function (i.e., the difference between the constrained and unconstrained values of the pseudo likelihood) is discontinuous in the structural parameters β . Discontinuities arise when applying I-I to discrete choice models because any simulated choice $\tilde{y}_{it}^m(\beta)$ is discontinuous in β (holding fixed the set of random draws used to generate simulated data from the structural model). To permit use of fast gradient-based numerical optimization algorithms that require a differentiable object function, we follow ASV and use a modification to I-I proposed in Keane and Smith (2003). Suppose that the simulated value of a binary variable $\tilde{y}_{it}^m$ equals 1 if a simulated latent utility $\tilde{u}_{it}^m(\beta)$ is positive and equals 0 otherwise. Rather than using $\tilde{y}_{it}^m(\beta)$ when computing $\tilde{\theta}(\beta)$, we use a continuous function $g(\tilde{u}_{it}^m(\beta); \lambda)$ of the latent utility. The function g is chosen so that as a smoothing parameter λ goes to 0, $g(\tilde{u}_{it}^m(\beta); \lambda)$ converges to $\tilde{y}_{it}^m(\beta)$. Letting λ go to 0 as the observed sample size goes to infinity ensures that $\tilde{\theta}(\beta_0)$ converges to θ_0 , thereby delivering consistency of the I-I estimator of β_0 . Our choice of g is

$$g(\tilde{u}_{it}^m(\beta); \lambda) = \frac{\exp(\tilde{u}_{it}^m(\beta)/\lambda)}{1 + \exp(\tilde{u}_{it}^m(\beta)/\lambda)}.$$

Because the latent utility is a continuous and smooth function of the structural parameters β , g is a smooth function of β . Moreover, as λ goes to 0, g goes to 1 if the latent utility is positive and to 0 otherwise.

When the structural model contains additional variables that depend on current and lagged values of indicator variables $\tilde{y}_{it}^m$, these additional variables will also be discontinuous in β . In our structural model, for instance, variables such as marriage duration and number of children depend on the history of indicator variables such as marital status and child birth. Since marriage duration and the number of children are discontinuous in β , they also contribute to creating a discontinuous objective function in the estimation process. Our smoothing strategy (which will be discussed in more detail in an Appendix to be provided in a future draft) ensures that all these variables will also be continuous in β , provided that they depend continuously on $\tilde{y}_{it}^m$. In other words, replacing the indicator functions by their continuous approximations $g(\tilde{u}_{it}^m(\beta); \lambda)$ ensures that all other variables that depend on β through $g(\tilde{u}_{it}^m(\beta); \lambda)$ are continuous.

As discussed in ASV, care must be taken in choosing λ , because approximation error in indicator functions for a particular year accumulate in the approximate functions that depend on those indicators. One needs to search for a combination of the smoothing parameter λ and the number of simulations M that generates sufficient smoothness in the objective function, while keeping

$t_i > t_{i \min}$, where $t_{i \min}$ is first observation on i . These approaches would involve adding a large number of parameters to the model, because $t_{i \min}$ varies substantially and the model involves multiple equations with a large number of endogenous state variables.

bias small and computation time manageable. The larger these parameters are, the smoother the objective function will be, but large values of λ introduce bias and large values of M increase computation time. Based upon simulation experiments, we chose a small value of λ , .05, which is large enough to smooth the objective surface sufficiently given our choice of 50 for M . Our simulation experiments indicate that the associated bias in the estimates is small for almost all of our parameters.

4.1 Auxiliary Model

As already mentioned, the auxiliary model consists of eight separate “blocks”, indexed by j , each tailored to identify a particular subset of parameters of the structural model. The criterion function, which is optimized in estimation, is $\sum_{j=1}^J \mathcal{L}_j$, where $\mathcal{L}_j$ is the likelihood function associated with block j . Each of the eight blocks consists of a system of seemingly unrelated regressions (SUR) with q_j equations and k_j covariates that are common to all q_j equations. Each block may be written as¹²

$$(29) \quad Y_{it}^{(j)} = Z_{it}^{(j)} \Pi^{(j)} + u_{it}^{(j)}; \quad u_{it}^{(j)} \sim N(0, \Sigma^{(j)}); \quad u_{it}^{(j)} \text{ i.i.d. over } i \text{ and } t.$$

Our choice of auxiliary model is motivated by the following principles. First, we use a rich auxiliary model rather than focus on a few features of the data. We do this because our model is intended to explain both contemporaneous and dynamic interrelationships among several labor market and marriage market variables. Given our objective, it makes more sense to use a rich auxiliary model that can capture these relationships rather than focus on a few features of the data. Second, each block uses a common set of right-hand-side variables to avoid having to iterate between Π and Σ to maximize the likelihood function. Third, by allowing for various separate blocks, we can tailor the right-hand-side variables to a particular subset of dependent variables. Some of the blocks of the auxiliary model are estimated using a particular subsample of observations, where the subsamples are typically defined based on the realization of a particular variable that is endogenous to the model, such as MAR_{it} and E_{it} . We next define the data vectors $Y_{it}^{(j)}$ and data matrices $Z_{it}^{(j)}$ for the eight blocks.¹³

¹²Note that the assumption $u_{it}^{(j)} \sim N(0, \Sigma^{(j)})$ with $u_{it}^{(j)}$ i.i.d. over i and t is false for several reasons, including the fact that $Y_{it}^{(j)}$ typically contains binary variables. The use of a misspecified likelihood affects efficiency but not consistency.

¹³As already noted, in our analysis we remove economy-wide year effects by first regressing measured wages, hours, and earnings on X_{it} and a set of year dummies. We then work with variables $\widetilde{wage}_{it}^*$, $\widetilde{hours}_{it}^*$, and $\widetilde{earnings}_{it}^*$, which refer to the residuals from such regressions. The vector X_{it} includes F_i , $EDUC_i$, a cubic polynomial in PE_{it} , a cubic polynomial in PE_{it} interacted with F_i , and the interaction of F_i with a quadratic time trend. An inconsistency arises because we do not include trends from employment in the structural model. Also, sample selection bias will arise in the estimates of the experience profiles of wages and hours. ASV estimate the PE_{it} profiles jointly with the other parameters.

4.1.1 Blocks 1-3

The first three blocks of the auxiliary model provide information used to identify the parameters of the labor market equations (i.e. employment transitions, wages, hours, and earnings). For the identification of this part of the model, we rely on information on sample members only. Block 1 focuses on sample members (male or female) who were employed in $t - 1$, while block 2 focuses on sample members who were not employed in $t - 1$. Block 3 provides information used in the estimation of the initial conditions of employment status and the autoregressive wage component, and is estimated on a subsample of sample members with relatively low age.

Block 1 :

$$\begin{aligned}
 (30) \quad Y_{it}^{(1)} &= [E_{it}, \widetilde{wage}_{it}^*, \widetilde{hours}_{it}^*, \widetilde{earnings}_{it}^*, \ln(1 + \widetilde{wage}_{it}^{*2})]' \\
 Z_{it}^{(1)} &= [constant, PE_{i,t-1}, PE_{i,t-1}^2, F_i, EDUC_i, E_{i,t-2}, \widetilde{wage}_{i,t-1}^*, \widetilde{wage}_{i,t-2}^*, \widetilde{hours}_{i,t-1}^*, \\
 &\quad \widetilde{hours}_{i,t-2}^*, \widetilde{earnings}_{i,t-1}^*, \widetilde{earnings}_{i,t-2}^*, \widetilde{wage}_{i,t-1}^* PE_{i,t-1}, \widetilde{wage}_{i,t-1}^* PE_{i,t-1}^2, \\
 &\quad \widetilde{wage}_{i,t-1}^* E_{i,t-2}, \widetilde{hours}_{i,t-1}^* E_{i,t-2}, MAR_{it}, F_i MAR_{it}, F_i CH01_{it}]'
 \end{aligned}$$

As already mentioned, Block 1 is estimated on the subset of sample members who were employed in $t-1$ (i.e. $E_{i,t-1} = 1$). Since the purpose of the structural model is to explain the behavior of the various dependent variables, including the labor market variables, block 1 includes each of the labor market variables in the structural model as a dependent variable. This accounts for the first four equations in the SUR system above. The fifth equation has $\ln(1 + \widetilde{wage}_{it}^{*2})$ as the dependent variable, which helps identify parameters of the model that influence the level and change in the variance of wages and earnings over time. It is also natural to use the explanatory variables in the structural model as right-hand-side variables in the auxiliary model. This accounts for the presence of $PE_{i,t-1}$, $PE_{i,t-1}^2$, F_i , and $EDUC_i$. Since the model is dynamic and includes state dependence terms in most equations, we include two lags of each of the dependent variables for the first four equations (except for $E_{i,t-1}$, which always equals 1 in this subsample). The lags help distinguish between state dependence and heterogeneity. We include the interaction terms $\widetilde{wage}_{i,t-1}^* PE_{i,t-1}$ and $\widetilde{wage}_{i,t-1}^* PE_{i,t-1}^2$ to capture changes with potential experience in the degree of persistence in wages. Finally, the terms MAR_{it} , $F_i MAR_{it}$, and $F_i CH01_{it}$ capture effects of marital status and of the presence of young children on the dependent variables.

Block 2 :

$$(31) \quad \begin{aligned} Y_{it}^{(2)} &= [E_{it}, \widetilde{wage}_{it}^*, \widetilde{hours}_{it}^*, \widetilde{earnings}_{it}^*]' \\ Z_{it}^{(2)} &= [constant, PE_{i,t-1}, PE_{i,t-1}^2, F_i, EDUC_i, E_{i,t-2}, \widetilde{wage}_{i,t-2}^*, \widetilde{hours}_{i,t-2}^*, \\ &\quad \widetilde{earnings}_{i,t-2}^*, \widetilde{wage}_{i,t-2}^* PE_{i,t-2}, MAR_{it}, F_i MAR_{it}, F_i CH01_{it}]' \end{aligned}$$

Block 2 is estimated on the subsample of sample members who were not employed in $t-1$ (i.e. $E_{i,t-1} = 0$). The vector of dependent variables is similar to that of block 1, except that $\ln(1 + \widetilde{wage}_{it}^{*2})$ is not included in this block. The rationale for the choice of variables to include in this block is similar to that of block 1.

Block 3 :

$$(32) \quad \begin{aligned} Y_{it}^{(3)} &= [E_{it}, \ln(1 + (\widetilde{wage}_{it}^*)^2)]' \\ Z_{it}^{(3)} &= [constant, DAGE26_{it}, DAGE27_{it}, DAGE28_{it}, DAGE29_{it}, F_i, EDUC_i]', \end{aligned}$$

where $DAGE26_{it}$ is a dummy that takes the value 1 when $AGE_{it} = 26$. [NOTE: We now draw $E_{i,1}$ jointly with $MAR_{i,1}$ and $NCH_{i,1}$.]

Block 3 is estimated on the subsample of sample members with low values of age (between 25 and 29) and provides information that is used in pinning down the employment status and the variance of the autoregressive component of wages at age 25.

4.1.2 Blocks 4-6

Blocks 4 to 6 provide information on the parameters of the marriage transition equations. Block 4 focuses on the *married-married* transition, while block 5 focuses on the *nonmarried-married* transition. Block 6 provides some complementary information.

Block 4 :

$$(33) \quad \begin{aligned} Y_{it}^{(4)} &= MAR_{it} \\ Z_{it}^{(4)} &= [constant, MDUR_{i,t-1}, AGE_F_{i,t-1}, AGE^2_F_{i,t-1}, AGE_M_{i,t-1}, AGE^2_M_{i,t-1}, \\ &\quad EDUC_F_{i,t-1}, EDUC_M_{i,t-1}, E_F_{i,t-1}, E_M_{i,t-1}, \\ &\quad \widetilde{wage_F}_{i,t-1}^*, \widetilde{wage_M}_{i,t-1}^*, (EDUC_i - EDUC_S_{i,t-1})^2, CH01_{i,t-1}]' \end{aligned}$$

Block 4 is estimated on the subsample of sample members who were married in $t-1$ (i.e. $MAR_{i,t-1} = 1$). This block has a single dependent variable, MAR_{it} . The set of right-hand-side

variables corresponds closely to the variables that determine the married-married transition in the structural model. Note that the variables are set up so that one set corresponds to the male and the other to the female (see discussion of the marriage transition equations in the structural model).

Block 5 :

$$\begin{aligned}
 Y_{it}^{(5)} &= MAR_{it} \\
 (34) \quad Z_{it}^{(5)} &= [constant, SDUR_{i,t-1}, AGE_F_{i,t-1}, AGE^2_F_{i,t-1}, AGE_M_{i,t-1}, AGE^2_M_{i,t-1}, \\
 &\quad EDUC_F_{i,t-1}, EDUC^2_F_{i,t-1}, EDUC_M_{i,t-1}, EDUC^2_M_{i,t-1}, \\
 &\quad E_F_{i,t-1}, E_M_{i,t-1}, \widetilde{wage_F}_{i,t-1}^*, \widetilde{wage_M}_{i,t-1}^*, CH01_{i,t-1}]'
 \end{aligned}$$

Block 5 is estimated on the subsample of sample members who were not married in $t - 1$ (i.e. $MAR_{i,t-1} = 0$). The dependent variable is MAR_{it} , and the vector of dependent variables is similar to that of block 4, except that it will not include observations corresponding to the spouse, since in this subsample ($MAR_{i,t-1} = 0$), there was no spouse at $t - 1$.

Block 6 :

$$\begin{aligned}
 Y_{it}^{(6)} &= MAR_{it} \\
 (35) \quad Z_{it}^{(6)} &= [constant, SDUR_{i,t-1}, AGE_F_{i,t-1}, AGE^2_F_{i,t-1}, AGE_M_{i,t-1}, AGE^2_M_{i,t-1}, \\
 &\quad EDUC_F_{i,t-1}, EDUC^2_F_{i,t-1}, EDUC_M_{i,t-1}, EDUC^2_M_{i,t-1}, \\
 &\quad E_F_{i,t-1}, E_M_{i,t-1}, \widetilde{wage_F}_{i,t-1}^*, \widetilde{wage_M}_{i,t-1}^*, CH01_{i,t-1}]'
 \end{aligned}$$

Block 6 is estimated using all observations on sample members, regardless of marital status. Note that block 6 and block 5 have the same $Y_{it}^{(j)}$ and $Z_{it}^{(j)}$, but are estimated on different subsamples. We include block 6 because it appears to provide some additional information that helps with the identification of the parameters of the marriage transition equations.

4.1.3 Blocks 7-8

Blocks 7-8 provide information about the dependence of the characteristics of spouses on the characteristics of sample members.

Blocks 7 – 8 :

$$\begin{aligned}
 (36) \quad Y_{it}^{(j)} &= [EDUC_S_{it}, AGE_S_{it}, \widetilde{wage_S}_{it}^*, \widetilde{hours_S}_{it}^*, \widetilde{earnings_S}_{it}^*, E_S_{it}]' \\
 Z_{it}^{(j)} &= [constant, F_i, EDUC_i, PE_{it}, PE_{it}^2, \widetilde{wage_it}^*, \widetilde{hours_it}^*, \widetilde{earnings_it}^*, E_{it}]'
 \end{aligned}$$

Block 7 is estimated on the subsample of sample members who are married at t . Block 8 is identical to Block 7, but is estimated on the subsample of sample members who are in their first year of marriage at t (i.e. $MAR_{it} = 1$ & $MAR_{i,t-1} = 0$).

4.2 Mechanics of Estimation

This section discusses a few additional details about the implementation of our estimation procedure. We begin with a discussion of the construction of the various subsamples that are used in the different blocks of the auxiliary model. As noted above, these subsamples are typically defined based on the realization of a particular variable that is endogenous to the model, such as MAR_{it} and E_{it} . Because of the smoothing procedure of discrete variables described in the previous section, the simulated values of variables MAR_{it} and E_{it} are continuous in the unit interval rather than binary (though most simulated values are close to either 0 or 1). As a result, the construction of the subsamples is done probabilistically, using probability weights. For example, for block 1, which restricts the sample to PSID sample members who are employed in $t - 1$ (i.e. $E_{i,t-1} = 1$), when computing $\hat{\Pi}^{(1)}$ in the SUR system $Y_{it}^{(1)} = Z_{it}^{(1)}\Pi^{(1)} + u_{it}^{(1)}$, we use all (i, t) observations in the simulation, but use the weighted variables $Y_{it}^{(1)} \cdot weight_{it}$ and $Z_{it}^{(1)} \cdot weight_{it}$, where $weight_{it} = E_{i,t-1}$ (where $E_{i,t-1}$ has been smoothed). For block 2, the weight used is $weight_{it} = 1 - E_{i,t-1}$. For block 8, the weight is $weight_{it} = MAR_{it}(1 - MAR_{i,t-1})$. Weights are defined in a similar fashion for the other blocks of the auxiliary model.¹⁴

Next, we note that, not surprisingly given the size and complexity of our models, the objective function displays multiple local optima with respect to some of the parameters. We experiment extensively with different starting values to make sure that we are finding the global optimum. We begin the process with estimates obtained from probit or regression models relating the dependent variable in each equation of the structural model to the observed variables in that equation, with the fixed heterogeneity components ignored. We refine our search by using grid evaluations, paying particular attention to the set of parameters that appeared most problematic, and by experimenting with smaller versions of our models to help us find good initial guesses, and then building up to more complex versions of the models. We view the results that we have obtained so far as provisional, not only because we impose a number of parameter restrictions that we will relax in a future draft, but also because more experimentation with starting values is needed.

Finally, we note that our chosen values of $\lambda = 0.05$ and $M = 50$ yield a smooth objective function that allows the use of fast gradient-based optimization algorithms with little evidence of

¹⁴We do not in fact use *all* (i, t) observations simulated, but rather those (i, t) that are available in the corresponding PSID sample, regardless of the value of the endogenous variables that are used to define the subsamples (such as MAR_{it} and E_{it}).

bias.¹⁵ The fact that we are iterating on a large number of parameters, the large size of some of the blocks in the auxiliary model, and the number of simulations make computation time-consuming even though we use a fast gradient-based optimization algorithm. To reduce estimation time, we exploit the highly parallelizable structure of our estimation methodology.¹⁶

4.3 Bootstrap Standard Error Estimation

We use a parametric bootstrap procedure to conduct inference. Given consistent estimates of the structural parameters, we repeatedly generate “artificial” observed data sets from the structural model, applying data availability rules that match the PSID sample. We treat each of the artificial data sets as if it was the PSID and apply our full estimation procedure, to obtain estimates of the parameters of the structural model for each such data set. The standard deviations of the parameter estimates across the data sets serve as our standard error estimates. The bootstrap procedure is very computationally intensive, so we use [300] bootstrap replications.

5 Model Estimates (Preliminary)

As already noted, estimation of our model presented in section 2, is very challenging. Our strategy has been to start with relatively simple versions of the model, and gradually build up towards richer specifications, while gaining experience with the practical details of estimation, including the specification of the auxiliary model. In fact, even the simplified versions of the model can be quite difficult to estimate. As part of the process of a gradual buildup towards richer specifications, we extensively experiment using Monte Carlo simulations. In particular, we simulate data from an estimated version of our model, treat the simulated data as if it were the true data, and then try to re-estimate the model as a way to check whether the auxiliary model provides sufficient information to recover the values of the structural parameters that were used to generate the “pseudo true” data. In other words, this procedure allows us to check whether our estimation procedure (including the specification of the auxiliary model) provides local identification of the structural parameters, as well as check for the presence of bias.

We have built up to and successfully estimated a version of the model that essentially imposes the following simplifying restrictions on the model presented in section 2:

- The factor loadings on η_i , the second source of permanent unobserved heterogeneity, are set

¹⁵We use a standard quasi-Newton algorithm with line search, which can additionally handle simple bounds on the parameter values. The algorithm approximates the (inverse) Hessian by the BFGS formula, and uses an active set strategy to account for the bounds. Gradients are computed by finite differences.

¹⁶Specifically, for a given value of the structural parameters, the $M = 50$ simulations required to evaluate the objective function are essentially independent and can be conducted simultaneously by 50 different processors. All programs are written in FORTRAN 90 and use the Message Passing Interface (MPI) for the parallelization.

to zero in all equations. (That is, there is a single factor, μ_i and $\mu_{_s i}$, capturing permanent unobserved heterogeneity.)

- The coefficients on wage variables $wage'_{it}$ and $wage'_{_s it}$ in equations (4) and (5) are set to zero. (That is, these variables do not affect the employment transition probabilities.)
- The effects of most variables that enter the marriage transition probabilities are restricted to be equal for males and females. For example, in the marriage continuation equation, $\gamma_{AGE_F}^{MM} = \gamma_{AGE_M}^{MM}$, $\gamma_{EDUC_F}^{MM} = \gamma_{EDUC_M}^{MM}$, $\gamma_{E_F}^{MM} = \gamma_{E_M}^{MM}$, $\gamma_{w_F}^{MM} = \gamma_{w_M}^{MM}$, and $\delta_{\mu_M}^{MM} = \delta_{\mu_F}^{MM}$.
- The marriage transition equations do not include measures of the disparity (such as the absolute difference in education) between the characteristics of potential marriage partners or spouses.
- In the hours equation, the effect of E_{it} is currently not allowed to vary by gender or marital status.
- In the fertility equations, the birth process currently depends only on marital status and age.

For this restricted model, simulations suggest that our estimates of most parameters are reasonably precise (we compute parametric bootstrap standard errors around the estimates). There is little evidence of bias. While we are currently working on enriching the model and estimating the gradually richer versions of the model, we provide a brief description here of our estimation results obtained with the restricted versions of the model successfully estimated thus far.

Our estimates of the restricted version of the model described above reveal the following: (i) Our estimates imply a strong degree of assortative mating on education, as well as on the autoregressive component of hourly wages. Assortative mating on the permanent unobserved component appears to be less important. (ii) The presence of young children has a very large effect on the probabilities of both marrying and of staying married. Young children also have a large effect on the probabilities that women are employed and stay employed. By contrast, education appears to play a relatively small role on the probabilities of marrying and of staying married, although it matters for employment transitions. (iii) Permanent unobserved heterogeneity (μ_i) plays a large role in basically all equations, although it appears to matter less for the continuation of an existing marriage.

Furthermore, simulations of the estimated model indicates that the model fits most of the variables in our PSID sample fairly well. In particular, data simulated from the estimated model matches reasonably well the means and variances of most variables in the model, as well as the evolution of key variables with age and/or experience. The data also match fairly well the strength

of the comovements with other variables, as measured by the coefficients obtained from regressions of key variables against other variables in the model, estimated on both the simulated data and the PSID data. One dimension where the current version of our estimated model is missing is in the distribution of work hours.

We will present the full estimation results from further expanded versions of the model in future drafts.

5.1 Using the Model

5.1.1 Impulse response functions

(Not available for this draft.)

- For single men and for women, at particular value of t , simulate effect of wage shocks, unemployment shocks, children, education, μ_i on
 - time path of expected prob. of marriage
 - characteristics of spouse conditional on marriage
 - time path of own earnings
 - time path of normalized family income
 - * decompose time path of normalized family income into effects on own earnings components, marriage, spouse’s earnings, etc.
- Similar simulations for married men and women
 - e.g. expected probability of marriage

5.1.2 Decomposition of the Variance of Lifetime Family Income per Adult Equivalent

(Not available for this draft.)

6 Conclusion

(Not available for this draft.)

7 References

Abowd, J.M. and Card, D.E. (1987). “Intertemporal labor supply and long-term employment contracts”, *American Economic Review*, 77(1), 50-68.

- Abowd, J.M. and Card, D.E. (1989). "On the covariance structure of hours and earnings changes", *Econometrica*, 57(2), 411-445.
- Altonji, Joseph, Anthony Smith, and Ivan Vidangos, "Modeling Earnings Dynamics," *Econometrica*, 81 (2013), 1395-1454.
- Altonji, J. G., A.P. Martins and A. Siow (2002). "Dynamic Factor Models of Wages, Hours, and Earnings", *Research in Economics* 56(1), 3-59.
- Altonji, J. G. and N. Williams (2005). "Do Wages Rise With Job Seniority? A Reassessment", *Industrial and Labor Relations Review*, 58(3), 370-397.
- Bagger, J., F. Fontaine, F. Postel-Vinay, and J.M. Robin (2013). "Tenure, Experience, Human Capital and Wages: A Tractable Equilibrium Search Model of Wage Dynamics", unpublished paper, University of Bristol.
- Baker, Michael and Gary Solon (2003) "Earnings Dynamics and Inequality Among Canadian Men, 1976-1992: Evidence from Longitudinal Income Tax Records", *Journal of Labor Economics* 21 (2003), 289–321.
- Barlevy, Gadi, (2008) "Identification of Search Models Using Record Statistics", *Review of Economic Studies*, 75(1), 29-64.
- Becker, Gary (1973). "A Theory of Marriage: Part I", *Journal of Political Economy* 81 (July 1973): 813–46.
- Becker, Gary (1974). "A Theory of Marriage: Part 2", *Journal of Political Economy* 82 (1974), s11-s26.
- Becker, G. S. (1981): "Altruism in the Family and Selfishness in the Market Place," *Economica*, 48(189), 1–15.
- Becker, Gary, Elisabeth Landes and Robert Michael (1977). "An Economic Analysis of Marital Instability", *Journal of Political Economy*, 85, 1141-1188.
- Bound, J., Brown, C., and Mathiowetz, N. (2001). "Measurement Error in Survey Data." in *Handbook of Econometrics*, V. 5, eds. E. E. Leamer and J. J. Heckman, pp 3705-3843.
- Browning, Martin, Pierre-André Chiappori, and Yoram Weiss (2011). "Family Economics." Cambridge University Press, forthcoming.
- Buchinsky, M., Fougère, D., Kramarz, F. and Tchernis, R. (2010). "Interfirm Mobility, Wages, and the Returns to Seniority and Experience in the U.S." *Review of Economic Studies*, 77(3) (July): 972-1001.
- Burdett, K., and M. G. Coles (1997): "Marriage and Class," *Quarterly Journal of Economics*, 112(1), 141–168.
- Burdett, Kenneth, and Coles, Melvyn (1999). "Long-Term Partnership Formation: Marriage and Employment." *Economics Journal* 109 (June): F307–F334.

- Chiappori, P.-A., M. Iyigun, and Y. Weiss (2009). “Investment in Schooling and the Marriage Market,” *American Economic Review*, vol. 99(5), pages 1689-1713, December.
- Chiappori, P.-A., and S. Oreffice (2008): “Birth Control and Female Empowerment: An Equilibrium Analysis,” *Journal of Political Economy*, 116(1), 113–140.
- Chiappori, P.-A., and Y. Weiss (2006): “Divorce, Remarriage, and Welfare: A General Equilibrium Approach,” *Journal of the European Economic Association*, 4(2-3), 415–426.
- Chiappori, P.-A., and Y. Weiss (2007): “Divorce, Remarriage, and Child Support,” *Journal of Labor Economics*, 25(1), 37–74.
- Choo, Eugene and Aloysius Siow (2006): “Who Marries Whom and Why”, *Journal of Political Economy*, Vol. 114, No. 1, pp. 175-201.
- Collins, E., and McNamara, J. “The Job Search Problem as an Employer-Candidate Game.” *Journal of Applied Probability* 28 (October 1990): 815–27.
- DeBacker, J., B. Heim, V. Panousi, S. Rammath, and I. Vidangos (2013). “Rising Inequality: Transitory or Persistent? New Evidence from a Panel of U.S. Tax Returns”, *Brookings Papers on Economic Activity*, Spring, pp. 67-122.
- Fernández, R. , N. Guner, and J. Knowles (2005): “Love and Money: A Theoretical and Empirical Analysis of Household Sorting and Inequality,” *Quarterly Journal of Economics*, 120(1), 273–344.
- Fernández, Raquel, and Richard Rogerson (2001). “Sorting and Long-Run Inequality”, *Quarterly Journal of Economics* , Vol. 116, No. 4 (Nov., 2001) , pp. 1305-1341.
- Gottschalk, P., and R. Moffitt (2011). “Trends in the Transitory Variance of Male Earnings in the U.S.: 1970-2004”, draft.
- Gourieroux, C., Monfort, A., and Renault, E. (1993). “Indirect Inference”, *Journal of Applied Econometrics* 8, S85-S118.
- Guvenen, Fatih, “An Empirical Investigation of Labor Income Processes,” *Review of Economic Dynamics*, 12 (2009), 58-79.
- Haider, S.J. (2001). “Earnings Instability and Earnings Inequality of Males in the United States: 1967-1991”, *Journal of Labor Economics*, 19(4), 799-836.
- Heckman J. (1981). “The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process.” In *Structural Analysis of Discrete Data with Econometric Applications*, Manski CF, McFadden D (eds). MIT Press: Cambridge, MA; 179–195.
- Jacquemet N., Robin J-M. (2012) “Assortative Matching and Search with Labor Supply and Home Production”. CeMMAP WP n°07-13
- Keane, M. and Smith Jr., A . A. (2003). “Generalized Indirect Inference for Discrete Choice

Models”, unpublished manuscript, Yale University.

Kremer, M. (1997). “How Much Does Sorting Increase Inequality?,” *Quarterly Journal of Economics*, MIT Press, vol. 112(1), pages 115-39, February.

Low, H., Meghir, C., and Pistaferri, L. (2010). “Wage Risk and Employment Risk over the Life Cycle”, *American Economic Review*, 100(4): 1432–67.

Meghir, C. and Pistaferri, L. (2004). “Income variance dynamics and heterogeneity”, *Econometrica*, 72(1), 1-32.

Mortensen, Dale T. (1998), ‘Matching: Finding a Partner for Life or Otherwise’, *American Journal of Sociology*, 94 , S215-S240

Shimer, R., and L. Smith (2000): “Assortative Matching and Search,” *Econometrica*, 68(2), 343–369.

Siow, Aloysius (2009): “Testing Becker’s Theory of Positive Assortative Matching,” University of Toronto WP, (356).

Weiss, Yoram, "The formation and dissolution of families: Why marry? Who marries whom? And what happens upon divorce", Handbook of Population and Family Economics, M. R. Rosenzweig & Stark, O. (ed.), 1997.

Wong, L. Y. (2003): “Structural Estimation of Marriage Models,” *Journal of Labor Economics*, 21(3), 699–727.

Wooldridge, J. (2005). “Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity”, *Journal of Applied Econometrics*, 20(1), 39-54.

Table I
Characteristics of Sample Members and their Spouses in PSID Sample

Variable						
<i>PSID Sample Members</i>						
	Male			Female		
	Obs.	Mean	St. Dev.	Obs.	Mean	St. Dev.
Education (years)	29,019	13.411	2.409	30,082	13.053	2.167
Age	29,019	39.614	9.599	30,082	38.869	9.222
Potential Experience	29,019	20.115	9.929	30,082	19.756	9.624
Marriage (M_t)	29,019	0.817	0.387	30,082	0.753	0.431
Marriage Duration	28,427	12.894	11.275	29,714	12.353	11.534
Employment	29,014	0.930	0.256	30,080	0.704	0.457
$E_t \mid E_{t-1} = 1$	24,797	0.965	0.184	19,302	0.905	0.293
$E_t \mid E_{t-1} = 0$	1,726	0.440	0.497	8,129	0.241	0.428
wage* _t $E_t=1$ ^(a)	21,618	0.000	0.426	16,577	-0.001	0.433
hours* _t ^(a)	28,532	-0.009	0.540	29,606	-0.008	0.963
earn* _t ^(a)	28,532	0.018	0.905	29,606	-0.009	1.312
Number of children	28,199	1.094	1.155	29,568	1.130	1.176
<i>Sample Members' Spouses</i>						
	Male			Female		
	Obs.	Mean	St. Dev.	Obs.	Mean	St. Dev.
Education (years)	22,559	13.327	2.461	23,631	13.028	2.054
Age	22,627	41.256	9.352	23,696	38.075	9.386
Potential Experience	22,627	21.860	9.692	23,698	19.021	9.687
Employment	22,577	0.943	0.233	23,678	0.670	0.470
$E_t \mid E_{t-1} = 1$	19,064	0.971	0.167	14,107	0.893	0.309
$E_t \mid E_{t-1} = 0$	1,028	0.427	0.495	7,040	0.225	0.418
wage* _t $E_t=1$ ^(a)	16,087	0.001	0.455	11,676	0.002	0.460
hours* _t ^(a)	22,304	-0.001	0.535	23,386	0.002	1.003
earn* _t ^(a)	22,304	0.000	0.973	23,386	0.010	1.365

The table presents descriptive statistics for key variables in PSID sample.

^(a) Variables $wage^*_{it}$, $hours^*_{it}$, and $earn^*_{it}$ are residuals from regressions of measured wages, hours, and earnings on gender, education, a cubic polynomial in potential experience, interactions of the polynomial with gender, the interaction of gender with a quadratic time trend, and a set of year dummies. The reference year is 1996.