Quantitative Economics 3 (2012), 471–525 1759-7331/20120471

Inequality and the life cycle

Greg Kaplan
Department of Economics, University of Pennsylvania

I structurally estimate an incomplete markets life-cycle model with endogenous labor supply using data on the joint distribution of wages, hours, and consumption. The model is successful at matching the evolution of both the first and second moments of the data over the life cycle. The key challenge for the model is to generate declining inequality in annual hours worked over the first half of the working life, while respecting the constraints imposed by the data on consumption and wages. I argue that this is a robust feature of the data on life-cycle labor supply that is strongly at odds with the intratemporal first-order condition for labor. Allowing for a realistic degree of involuntary unemployment, coupled with preferences that feature nonseparability in the disutility of the extensive and intensive margins of hours worked, allows the model to overcome this challenge. The results imply that labor market frictions are important in jointly accounting for observed cross-sectional inequality in labor supply and consumption, and may have quantitative relevance for analyses that exploit the intratemporal first-order condition for labor.

Keywords. Inequality, life cycle, hours worked, intensive and extensive labor supply, structural estimation, precautionary savings.

JEL classification. C13, D21, E21, E24, J22.

1. Introduction

Can endogenous labor supply choices in the presence of incomplete consumption insurance account for the patterns of inequality in consumption and hours over the life cycle? With regards to consumption, a large and growing literature has been broadly successful in accounting for both the life-cycle mean and the life-cycle variance.[1] This literature is motivated by two observations. First, life-cycle data on consumption are informative about the nature of risks that households face and the degree of insurance that is afforded by financial markets or other informal mechanisms.[2] Second, consumption is a direct measure of welfare. Hence understanding how it differs by age and in the cross section is an important goal for economic research.

Greg Kaplan: gkaplan@sas.upenn.edu
I am grateful for comments from Dirk Krueger, Tom Sargent, Sam Schulhofer-Wohl, Kjetil Storesletten, and Gianluca Violante, as well as participants at numerous seminars and four anonymous referees. All errors are my own.

[1]See, for example, Deaton and Paxson (1994), Blundell, Browning, and Meghir (1994), Attanasio, Banks, Meghir, and Weber (1999), Gourinchas and Parker (2002), Storesletten, Telmer, and Yaron (2004), Aguiar and Hurst (2007, 2008).

[2]See for example Blundell, Pistaferri, and Preston (2008) and Kaplan and Violante (2010).

DOI: 10.3982/QE200

However, the analogous literature has not successfully confronted the data on inequality in labor supply over the life cycle. Yet understanding cross-sectional differences in hours worked is an equally important research goal for the same two reasons. First, in addition to being informative about risk and insurance markets, labor supply data can help in understanding the structure of labor markets. Second, labor supply is also an argument of utility and hence is directly useful for measuring welfare. Thus there remains an important gap in our understanding of life-cycle inequality.

In models with incomplete consumption insurance and frictionless labor supply decisions, the difficulty in simultaneously accounting for the cross-sectional data on consumption and hours is due to an inconsistency between the intratemporal first-order condition for labor and one of the most robust life-cycle facts about the distribution of labor supply: hours inequality declines sharply over the first 15 years in the labor market. I argue that a general feature of this class of models is that they cannot generate declining inequality in annual hours, while still respecting the restrictions imposed by the data on wages and consumption.[3] Understanding this discrepancy between model and data is important because it undermines a trade-off that is at the heart of the predictions of numerous economic models: equating wages to the marginal rate of substitution between consumption and hours. Quantitatively, this inability of standard models to fit the data holds whether parameters are estimated to match only life-cycle means or whether life-cycle variances are also explicitly targeted.

In this paper, I structurally estimate an incomplete markets model using data on the joint distribution of both consumption and hours, and show that with a small modification to the standard model, it is possible to account for the key features of the life-cycle data.[4] Standard models fail in this respect because when the data are viewed through the lens of the intratemporal first-order condition for labor, there appears to be an idiosyncratic wedge whose cross-sectional variance declines sharply with age. The key ingredient that generates this feature in my model is involuntary unemployment. In the model, involuntary unemployment acts as shocks to individuals' endowments of time available for work. For such shocks to have an effect on labor supply, some degree of nonseparability between the disutility of the extensive and intensive margins of work is necessary: individuals care not only about the total number of hours that they work in a given year, but also about how those hours are spread across periods of employment and unemployment. This nonseparability converts these unemployment shocks into an idiosyncratic wedge in the intratemporal first-order condition. Since the incidence of unemployment declines sharply with age, the cross-sectional variance of this wedge also declines with age, generating a wedge with exactly the property that is missing from standard models. The extent of the nonseparability, along with the other model

[3]This result applies in a broad class of models. It holds for most reasonable preference formulations and does not depend on the statistical process governing wages. See Section 6 and Appendix H for further details. Storesletten, Telmer, and Yaron (2001) provided a related argument in a model with complete consumption insurance.

[4]The focus of the paper is on *residual* inequality, that is, cross-sectional differences that exist once cross-sectional differences in education, race, and other observable characteristics have been controlled for. This paper focuses exclusively on the labor supply of males.

parameters, is estimated using Simulated Method of Moments (SMM). Formal tests of overidentifying restrictions are constructed and are not rejected by the data.

The resulting model shares features with search models in that it emphasizes unemployment and acknowledges that part of the observed distribution of hours is due to constraints rather than choices. But, whereas most search models have little to say about consumption and wealth, my model is ultimately an incomplete markets models with endogenous labor supply. Hence it can speak to the data on consumption and wealth, as well as retain an active intensive labor supply margin, which I show to be important. In the model, some of the cross-sectional dispersion in annual hours worked arises from cross-sectional differences in choices, while some arises from cross-sectional differences in opportunities for work. Both factors contribute to the decline in dispersion at young ages.

I provide substantial evidence on the distribution of unemployment and hours over the life cycle to support the labor supply model that I adopt: I show that the decline in hours inequality is predominantly residual in nature, that is, it is not driven by observable characteristics. I show that the decline takes place both in usual hours worked in a given week as well as in the total number of weeks worked per year. I show that the decline is a phenomenon that lies exclusively in the bottom half of the hours distribution. Finally, I show that the fraction of the population who experience an unemployment spell during the year declines sharply with age, but that the distribution of time spent unemployed, conditional on it being positive, does not vary with age. These last two features of the data are what disciplines the amount of unemployment risk in the model.

The early literature on life-cycle labor supply focused on rationalizing the hump-shaped profile in mean hours over the life cycle and almost exclusively viewed the data through the lens of a risk-neutral agent, where consumption and earnings are equal.[5] This is limiting because not only is it difficult to speak to data on inequality, it implies that two separate theories are required to understand life-cycle consumption and life-cycle labor supply, neither of which can be held accountable to the joint restrictions that are imposed by the data. Almost all of the vast literature on labor market search in a life-cycle setting falls into this category.[6]

However, there are important reasons for studying the life-cycle distribution of labor supply jointly with consumption. The trade-off between consumption smoothing and/or redistribution, on one hand, and the efficient intertemporal allocation of labor, on the other hand, is at the heart of a number of important economic questions. Salient examples include the design of optimal income taxation, measuring the welfare implications of changes in technology, and the role of technology shocks for business cycle fluctuations.[7]

[5]This literature dates back to Heckman (1974), Heckman and MaCurdy (1980), and MaCurdy (1981).

[6]For a recent example, see Menzio, Telyuokva, and Visschers (2010). There are some exceptions: examples of papers that can speak to the life-cycle data on both consumption and labor supply include Pijoan-Mas (2006), Erosa, Fuster, and Kambourov (2011), Low, Meghir, and Pistaferri (2009), and Lise (2010). However, none of these papers estimates structural parameters from the joint restrictions imposed by consumption and hours.

[7]The intratemporal first-order condition is at the heart of the economic mechanism in each of these literatures, thus confronting this trade-off with microdata is of first-order importance. Examples of its impor-

The few existing models that relax risk neutrality have either not confronted or not been able to explain the declining profile for variance of hours early in the life cycle. Two of these papers that warrant further mention are Heathcote, Storesletten, and Violante (2009, 2010a). Heathcote, Storesletten, and Violante (2010a) used a similar model to understand low frequency time trends in inequality in the United States. That paper does not obtain structural estimates and is not consistent with the life-cycle data on labor supply. Hence both the goals and the methodology are very different from this paper.

Heathcote, Storesletten, and Violante (2009) did use structural estimation to confront a set of facts similar to what I study in this paper. However, there are two important differences between the model in Heathcote, Storesletten, and Violante (2009) and the model in this paper, both of which are crucial for being able to account for the life-cycle distribution of hours. The first difference is that there is no role for noncontingent nonhuman wealth in the partial insurance economy in Heathcote, Storesletten, and Violante (2009). In this paper, however, nonhuman wealth is the key ingredient needed to match the data on mean hours and is the main mechanism that allows households to substitute labor supply across time. The second difference is that Heathcote, Storesletten, and Violante (2009) used a perpetual-youth model, rather than a life-cycle model. Hence there are no horizon or retirement effects, which are the essence of a life cycle.

This paper is also related to the literature on structural estimation with incomplete markets.[8] The current paper goes beyond the scope of this existing work by addressing a wider set of facts that include information on both consumption and hours. Similarly, there is a parallel literature on the structural estimation of search models that focuses on estimating parameters of the process governing labor market outcomes, but ignores consumption. It is the joint restrictions on these variables that pose the biggest challenge for standard models.

Outline of the paper. In Section 2, I outline a benchmark model without involuntary unemployment. In Sections 3 and 4, I then discuss the estimation strategy and show that the benchmark model is successful when estimated using only life-cycle means. Section 5 shows that the same model fails when confronted with data on life-cycle variances, whether or not these are included as estimation targets. In Section 6, I use the first-order condition (FOC) for labor to illustrate the sources of this failure. Section 7 provides further evidence on the distribution of hours over the life cycle, including data on unemployment spells. In Section 8, I discuss ways to incorporate a role for unemployment into the benchmark model, and in Section 9, I show that estimation of the resulting model is successful. In Section 10, I evaluate the fit of the model along some additional dimensions and Section 11 concludes. There are many details and sensitivity analyses that need to be addressed in a large-scale structural estimation such as this. I have relegated many of these details to the Appendices in an attempt to improve the flow of information and to communicate better the main points of the paper.

tance include Heathcote, Storesletten, and Violante (2010a) for measuring welfare, Heathcote, Storesletten, and Violante (2010b) for optimal taxation in the Ramsey tradition, Kocherlakota (2010) for optimal taxation in the Mirlees tradition, and the vast literature on Real Business Cycle (RBC) models for business cycle fluctuations.

[8] Important related papers in this category include Gourinchas and Parker (2002), Imai and Keane (2004), French (2005), and Guvenen and Smith (2009).

2. A benchmark model

In this section, I describe a standard incomplete markets life-cycle economy, extended to include a nontrivial labor supply decision. Each household consists of a single worker with a fixed time endowment. In each year, the worker chooses what fraction of his time endowment to work and earns an hourly wage equal to his individual labor productivity. Individual labor productivity follows an exogenous stochastic process.

Demographics. The economy is populated with a continuum of households, indexed by i. Agents work until age T^{ret}, at which time they enter retirement. The unconditional probability of surviving until age t is denoted by S_t. I assume that $S_t = 1$ for the first $T^{\text{ret}} - 1$ periods, so that there is no chance of dying before retirement. After retirement, $S_t < 1$ and all agents die by age T with certainty. Households are assumed to have no bequest motive. To focus solely on wage uncertainty, I assume that there exist perfect annuity markets so that households are completely insured against survival risk. The model period is 1 year, to be consistent with the data on labor earnings and total hours worked in the Panel Study of Income Dynamics (PSID). For simplicity, I do not attempt to model household size or composition.[9]

Preferences. Households have time-separable expected utility preferences over annual consumption c_{it} and annual hours worked H_{it} given by

$$\mathbb{E}\sum_{t=1}^{T}\beta^{t-1}S_t\left[\frac{c_{it}^{1-\gamma}}{1-\gamma} - \varphi_i\frac{H_{it}^{1+\sigma}}{1+\sigma}\right].$$

In this specification, γ is the coefficient of relative risk aversion and $1/\sigma$ is the Frisch elasticity of labor supply.[10] I allow for fixed heterogeneity in households' distaste for working relative to consumption (φ_i).[11]

Technology. Each household consists of a single worker who can choose to devote a fraction of his time endowment (normalized to 1) to labor market activities. The worker

[9]There is substantial evidence that changes in household composition due to marriage and fertility may play an important role in determining the evolution of consumption, labor supply, and wealth over the life-cycle (see, for example, Blundell, Browning, and Meghir (1994) or Attanasio et al. (1999)), and that joint labor supply decisions or issues surrounding intrahousehold allocations are important (see, for example, Heckman and MaCurdy (1980) or Attanasio, Low, and Sánchez-Marcos (2005)). However, rather than attempting to explicitly incorporate such features into the model, I instead focus on the simplest possible model that can speak to the data and choose an appropriate sample: households with a male head who is the single primary earner. In Appendix C, I compare this sample with alternative choices and show that none of the main features of inequality is changed. Hence none of the arguments in the paper is sensitive to this restriction.

[10]For $\gamma \neq 1$, these preferences are not consistent with balanced growth. In particular, with $\gamma > 1$, they predict that the fraction of time devoted to labor will fall over time. In my benchmark sample of males, the average fraction of time spent working decreased by 1 percentage point between 1969 and 2004. None of the main points of the paper is affected by allowing for preferences that are nonseparable between consumption and hours. See Section 6 and Appendix H for further details.

[11]It is possible to allow for heterogeneity in any of the parameters of the utility function. However, my approach is to restrict heterogeneity to the minimum amount that is necessary to explain the key features of the data. Heterogeneity in the disutility of working turns out to be useful in this respect.

receives an individual specific wage w_{it} for each hour worked. Thus annual labor earnings are given by $y_{it} = w_{it} H_{it}$. Log wages follow an exogenous stochastic process that consists of four components,

$$\log w_{it} = \kappa_t + \alpha_i + z_{it} + \varepsilon_{it},$$

$$z_{it} = \rho z_{it-1} + \eta_{it},$$

where κ_t is a nonstochastic experience profile that is assumed to be the same for all individuals in the economy, α_i is an individual-specific fixed effect, z_{it} is an idiosyncratic (autoregressive) AR(1) persistent shock, and ε_{it} is an idiosyncratic independent and identically distributed (IID) shock.[12]

Asset markets. Households can hold quantities of a single risk-free security a_{it} that pays interest at a gross rate R. Holdings can be negative, subject to satisfying an exogenously specified borrowing constraint $\underline{a}$. I consider two alternatives for determining the tightness of borrowing limits: (i) excluding borrowing altogether, $\underline{a} = 0$; (ii) estimating $\underline{a}$ along with the other parameters of the model. Households are born with an initial wealth endowment a_{i0}, the distribution of which is calibrated to be consistent with data on the wealth of young households in the PSID.

Government. Households face a progressive tax on labor income, given by the function $\varsigma(y)$, a proportional tax on capital income τ_a, and a proportional consumption tax τ_c. The government also administers a progressive pay-as-you-go social security system, designed to mimic the redistribution implicit in the U.S. Social Security system. In the model, Social Security benefits are functions of the fixed component in wages, α_i.

Budget constraints. The budget constraint for a working age household is hence given by

$$(1 + \tau_c) c_{it} + a_{i,t+1} \leq R(1 - \tau_a) a_{it} + w_{it} H_{it} - \varsigma(w_{it} H_{it}),$$

$$H_{it} \leq 1,$$

$$a_{i,t+1} \geq \underline{a}.$$

In retirement, there is no labor supply decision. The budget constraints look the same except that earnings $w_{it} H_{it}$ are replaced with Social Security income and the interest rate is adjusted by the probability of survival to reflect the existence of perfect annuity markets.

[12]Each component is assumed to be drawn from a discretized normal distribution with zero mean. The variance of the IID shock is allowed to vary with age. All other variances are constant across ages. The initial draw of the persistent component z_{i0} is set to zero. The decision to adopt this particular statistical process for wages is motivated by the fact that it provides an excellent fit to the autocovariance structure of wages across ages. Appendix E contains a full description of the wage process, gives further details of the estimation approach, and proves that it is identified from available panel data on wages. In Section 6, I argue that none of the main points of the paper would be affected by adopting one of the other reasonable statistical processes that has been proposed in the literature, provided that it is consistent with the life-cycle patterns in the cross-sectional variance of wages in the data.

3. Taking the benchmark model to the data

In this section, I describe the most important of the choices that are needed so as to confront the model with data on wages, consumption, and hours. To conserve on space and progress quickly to the main results of the paper, I refer the reader to the Appendices for many of the details.

Data sources. Data on wages, hours, and wealth comes from the Panel Study of Income Dynamics (PSID). The long panel dimension of the PSID enables estimation of the exogenous stochastic process for wages in a first-stage process outside the model. The main limitation of the PSID is the lack of comprehensive data on consumption. Hence I use the Consumer Expenditure Survey (CEX), which is a cross-sectional survey, for data on consumption. The CEX also contains data on earnings and hours, from which I construct joint moments for consumption with hours and wages.

Sample selection. In both the PSID and the CEX the household is the relevant unit of analysis. I focus on a sample of households with one primary male earner. This sample is chosen since it most closely reflects the nature of the model described in Section 2. I focus on the working life only and do not attempt to match data on choices during retirement.[13]

Variable definitions. Hours are defined as total annual hours of work, constructed as actual weeks worked multiplied by usual hours worked per week. Labor market earnings are defined as wage and salary income from all jobs, including bonuses, tips, and overtime plus the labor part of income from self-employment. Wages are constructed as earnings divided by annual hours. For annual consumption, I use total nondurable expenditures.[14] An important consideration in confronting the model with data on consumption is how to account for differences in household composition across households. The choice of whether and how to make household expenditures equivalent can have a potentially large impact on the evolution of the mean and variance of consumption over the life cycle. Rather than enter into the debate about which equivalence scale

[13]Full details of each data set and the selection process can be found in Appendix A. The important features are as follows. For the PSID, a household is included in the sample if the head is a male aged between 20 and 60 with between 3 and 38 years of potential labor market experience (defined as age minus years of education minus 6), has nonmissing data on completed years of education, worked between 520 and 5200 annual hours, and whose nominal wage is at least as high as the corresponding minimum wage in that year. I use data from the 1970–2005 waves, covering earnings in 1969–2004. For the CEX, a household is included in the sample if it has a male head who satisfies the age, experience, hours, and minimum wage criteria described above. In addition, the household must have four completed quarterly interviews. A household is assigned to a given calender year if the fourth interview took place before April of the following year. I use data from 1980–2003. The selection of one primary male earner is imposed by dropping all household/year observations where the second earner has annual labor earnings equal to more than half of the head's annual labor earnings. In Appendix C.1, I report findings from two alternative samples: one that contains the full sample of households and one that is the complement of the main sample — married households with two earners. The broad patterns of first and second moments of the relevant variables are similar in these alternative samples. Hence none of the main points of the paper would be affected by using different selection criteria or focusing on a broader sample.

[14]Appendix C.4 reports how the data are affected by using broader measures of consumption that include expenditures on durables or imputed services from durables.

is the most appropriate, I conduct all the analysis using the OECD-modified equivalence scale and refer the reader to Appendix C.5 for a comparison with other reasonable equivalence scales.[15]

Definition of life cycle. I use potential labor market experience, defined as age minus years of education minus 6, as a measure of the life cycle and focus on the range of potential experience from 3 to 38.[16] All major features of the data are unaffected by using age, rather than experience, as the life-cycle variable.[17]

Measurement error. I allow for the possibility that log hours, log earnings, and log consumption are contaminated with classical measurement error. Log wages hence contain measurement error from both hours and earnings. When estimating the model using first moments only, I use an external estimate of measurement error in wages. When estimating the model with data also on second moments, I estimate measurement error in earnings and hours directly, generating an implicit estimate for measurement error in wages.

Construction of moments. The focus of the paper is on residual inequality: cross-sectional variation in consumption and labor supply that is not accounted for by fixed individual differences in observed characteristics. To extract the residual component of each variable, I run a first-stage regression on a full set of experience dummies, four education dummies interacted with either year or cohort dummies (see below), and race dummies.[18] Mean life-cycle profiles are constructed as the estimated coefficients on the experience dummies in this regression. Second moments are constructed using the residuals from this regression.[19] All confidence intervals are calculated by bootstrap.

Year effects versus cohort effects. It is well known that an identification problem arises when trying to measure the evolution of variables over the life cycle: one cannot simultaneously allow for distinct life cycle, year, and cohort effects in a separable model.[20]

---

[15]None of the findings about which features of the model are important for matching the data is affected by the choice of equivalence scale. However, the resulting parameter estimates would be affected. This is primarily due to the view one takes about how steep is the rise in consumption inequality over the life cycle, a feature that is strongly influenced by the choice of equivalence scale. There are many different views in the literature; see, for example, Deaton and Paxson (1994), Slesnick and Ulker (2004), Heathcote, Storesletten, and Violante (2005), Guvenen (2007), and Primiceri and Van Rens (2009). A comparison of alternative definitions of consumption and equivalence scales can be found in Appendix C.5.

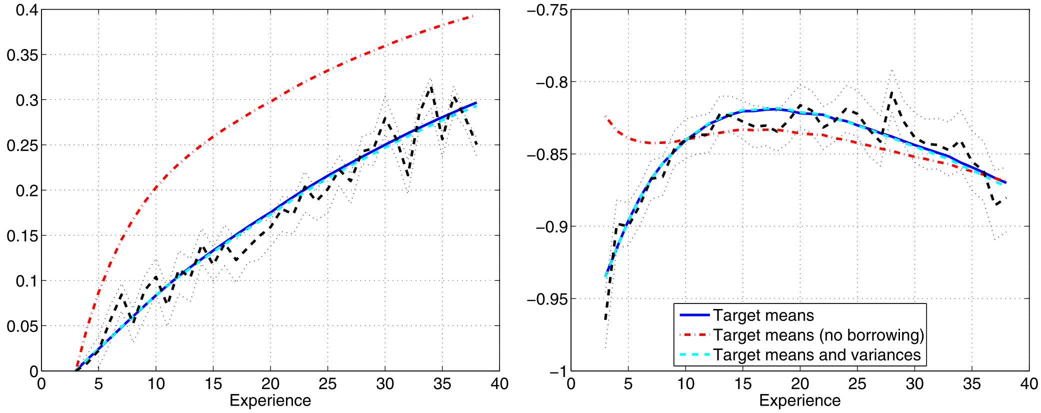
[16]I start at experience level 3 to minimize the impact of heterogeneity in the initial transition from education to the labor market. For example, the model does not include the possibility of gap years, travel, internships, or a long unemployment spell during the initial job search period.

[17]A comparison of the life-cycle moments using age versus potential experience can be found in Appendix C.2. The reason for using potential experience is to reflect the fact that in the model, agents are born at the time of entry into the labor market. Hence there is no distinction between two agents of different ages and different education levels with the same number of years of potential experience. Given that it is the accumulation of idiosyncratic productivity shocks that generates patterns of inequality in the model, the relevant life-cycle dimension on which to assess the model is potential experience, rather than age.

[18]The reason for focusing on residual inequality is that the model is stationary and does not allow for an education decision before entry to the labor market or for differential impacts of race on early human capital accumulation. Hence the model can be expected to explain only the variation in the data that remains after removing variation from these sources.

[19]Appendix B contains full details on the construction of the data that are ultimately used for estimation.

[20]See Hall (1971), Heckman and Robb (1985), Slesnick (2005), and Yang, Schulhofer-Wohl, Fu, and Land (2008) for alternative approaches to identification.



(a) Mean log consumption (b) Mean log hours

FIGURE 1. Model fit for first moments. The thick dashed line is the data; the thin dotted line is the 95% confidence interval for the data; the solid line is the fit of the model when only means are targeted; the dash-dot line is the fit of the model when only means are targeted and borrowing is not allowed; and the thin dashed line is the fit of the model when means and variances are jointly targeted.

Rather than enter the debate over whether it is more appropriate to control for year or cohort effects, I instead repeat all of the analysis twice: once controlling for cohort effects, once controlling for year effects. For brevity, I report only the results controlling for year effects in the main text. Analogous results, including all structural parameter estimates for the cohort view, can be found in Appendix I.[21]

Summary of key empirical patterns. The resulting life-cycle profiles for the key endogenous variables are presented and discussed alongside the estimation results in the sections that follow. Here I summarize the most important empirical patterns.

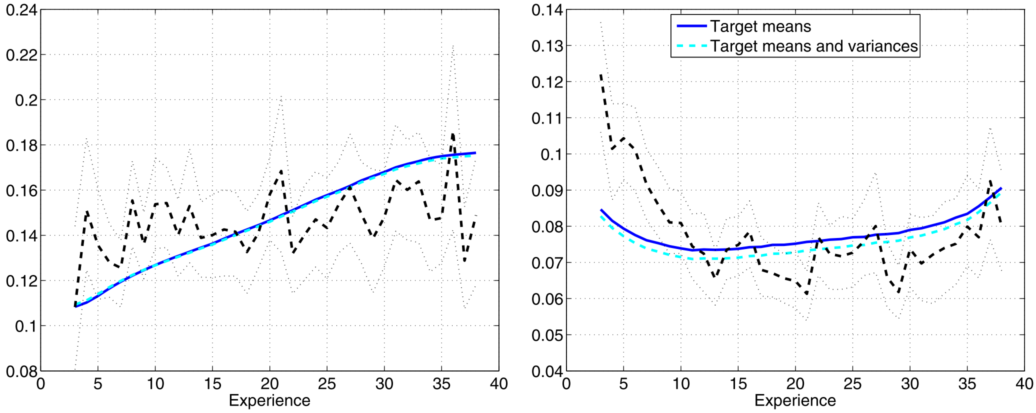
With regard to first moments (see Figure 1), there are two important features of the data:

- Mean log consumption increases roughly linearly during the working years.
- Mean log hours is inverse U-shaped. There is a sharp increase over the first 10 years, then a flattening, and eventually a decrease toward retirement.

With regard to second moments (see Figure 2), the important features are the following:

- The variance of log consumption increases by less than the variance of log wages. However, the magnitude of the increase depends on the equivalence scale. For the OECD-modified scale, the profile is almost flat.

[21]None of the main points of the paper is affected. There are, of course, differences in the estimated parameter values. The most important of these differences are a slightly higher estimate for the coefficient of relative risk aversion (γ) and a slightly lower estimate for the Frisch elasticity of labor supply ($\frac{1}{\sigma}$).



(a) Variance log consumption (b) Variance log hours

FIGURE 2. Model fit for second moments. The thick dashed line is data; the thin dotted line is the 95% confidence interval for the data; the solid line is the fit of the model when only means are targeted; and the dashed line is the fit of the model when means and variances are jointly targeted.

- The variance of log hours is strongly U-shaped. In particular, there is a sharp downward trend in cross-sectional inequality in hours worked over the first 15 years. This is the most important fact that I address in this paper. In Section 6, I explain why versions of the benchmark model have difficulty reproducing this downward trend. In Section 7, I investigate hours inequality more thoroughly and show that the U-shaped pattern is extremely robust: it is evident in other data sets, it is not driven by observable characteristics such as fertility and marriage, and it is not driven by compositional effects.

4. Successful estimation with first moments

The key parameters of the model are identified using only data on the evolution of mean consumption and mean hours over the life cycle. In this section, I exploit this to obtain parameter estimates and test the fit of the model with respect to these first moments.There are two reasons for starting by only focusing on first moments. First, I want to convey the notion that this model is successful in terms of matching life-cycle means. Second, I want to illustrate the crucial role of financial wealth in explaining the evolution of mean consumption and labor supply.

Estimation procedure. A full description and formal characterization of the estimation procedure can be found in Appendix G. Here I provide only an outline of the key features of the process. I adopt a simulated method of moments (SMM) estimator with a diagonal approximation to the optimal weighting matrix.[22] The targeted moments are

[22]Using an estimate of the optimal weighting matrix is useful since it permits a simple way to construct tests of overidentification. In similar contexts, it has been shown that the optimal weighting matrix may perform poorly in small samples and that an identity matrix may be preferred. See Altonji and Segal (1996). However, in this case, it is not obvious how such an identity matrix should be constructed, since different

mean residual log hours and log consumption at 36 experience levels (3–38). Since it is unreasonable to expect the model to match the *level* of consumption (due to the presence of unearned income, and the fact that the data are (a) made equivalent and (b) contain only nondurable expenditures), I normalize mean log consumption to zero at experience level 3. This generates a total of 71 target moments.

The estimated parameters are the coefficient of relative risk aversion (γ), the inverse of the Frisch elasticity (σ), the borrowing limit ($\underline{a}$), and the degree of cross-sectional variation in the fixed disutility of working (φ_i), which is expressed as a coefficient of variation.[23] The stochastic process for individual wages is estimated separately in a first stage and then fed into the model.[24] Confidence intervals are constructed by bootstrap to account for the additional estimation error induced by the first stage.

Internally calibrated parameters. Two additional constraints are imposed during the estimation. First, I vary the mean disutility of working ($E[\varphi_i]$) so that mean log hours across households of all experience levels is equal in the model and the data: 44% of the available time endowment. Second, I impose that the total amount of wealth held by households is consistent with the data. I do this by varying the discount factor β to match the 75th percentile of the distribution of wealth holdings for households entering retirement.[25] I do not treat β or $\mathbb{E}[\varphi_i]$ as parameters to be estimated. I treat the corresponding targets as nonrandom quantities.[26]

Externally calibrated parameters. I set the interest rate $R-1$ at 3%, the proportional tax on capital at 40%, and the consumption tax at 8%.[27] The progressive tax function for labor earnings is a smooth (differentiable) approximation to labor income taxes in the

moments are measured in different units. This difficulty becomes more apparent when second moments are also included as targets for estimation. The covariance matrix of target moments is approximated with a bootstrap estimator. Since the off-diagonal elements of this covariance matrix are very close to zero and imprecisely estimated, for simplicity I use only the diagonal elements of the matrix and set all others to zero. See Appendix G for details.

[23] I assume that φ_i is drawn from a log-normal distribution. Since the mean is calibrated internally, estimating the coefficient of variation is equivalent to estimating the variance, but is numerically more stable.

[24] The parameter estimates and the model fit for the stochastic wage process are described in Appendix E. The assumed process provides an excellent fit to the autocovariance structure of wages over the life cycle.

[25] Since wealth is the key feature of the model that allows households to smooth consumption and labor supply over time, it is crucial that the model implies a distribution of wealth holdings that is consistent with the data. See Storesletten, Telmer, and Yaron (2004) for a discussion of the importance of matching the wealth level for predictions about consumption inequality over the life cycle. Since it is well known that models with idiosyncratic productivity shocks as the only source of heterogeneity have difficulty in matching the upper tail of the wealth distribution, I consider it more important to target a feature of the distribution that is not too heavily biased by the right tail. Hence I focus on the 75th percentile at the point where average wealth holdings are greatest. The level of this target is 3.52 times average annual wages.

[26] An alternative approach would be to include β and $E[\varphi_i]$ as two additional parameters to be estimated and to include these two additional moments as targets. There are two reasons why I prefer to calibrate them. The first reason is computation. These two parameters are uniquely pinned down by these two moments. Hence simple nested nonlinear equation solvers can be used for calibration. This makes the nonlinear optimization process that is required for the other parameters simpler. The second reason is that by calibrating these parameters, the spirit of the exercise is clear: How well can the model match the life-cycle means and variances, *conditional* on being consistent with overall hours and wealth in the data?

[27] For the capital tax, I follow Domeij and Heathcote (2004). The consumption tax is based on evidence in McDaniel (2007).

TABLE 1. Parameter estimates.[a]

	(1)	(2)	(3)
Estimated parameters			
γ	1.60	1.86	1.65
	(1.22, 1.98)	(1.57, 2.68)	(1.36, 2.15)
σ	2.00	2.48	1.94
	(1.31, 3.13)	(1.37, 5.86)	(1.31, 2.67)
CV[φ]	1.27	0.98	1.16
	(0.65, 142.0)	(0.55, 3.40)	(0.72, 9.77)
$\underline{a}$	−0.30		−0.27
	(−6.90, −0.10)		(−0.63, 0.09)
$\sigma^2_{\epsilon,y}$			0.00
			(0.00, 0.05)
$\sigma^2_{\epsilon,h}$			0.03
			(0.00, 0.05)
$\sigma^2_{\epsilon,c}$			0.05
			(0.03, 0.08)
Test of overidentifying restrictions			
p-value	0.45	0.00	0.00
Calibrated parameters			
β	0.990	0.988	0.990
$E[\varphi]$	29.4	67.2	30.3
Targeted moments			
$E[\log c]$	X	X	X
$E[\log h]$	X	X	X
$V[\log c]$			X
$V[\log h]$			X

[a]95% bootstrap confidence intervals are given in parentheses. The borrowing limit $\underline{a}$ is expressed as a multiple of the average annualized hourly wage. $\sigma^2_{\epsilon,y}$ and $\sigma^2_{\epsilon,h}$ are estimates of classical measurement error in earnings and hours, respectively; when not estimated, they are set at 0.01.

United States, described in Appendix F.2. The calibration of Social Security payments is described in Appendix F.1. I estimate the distribution of initial wealth a_{i0} from the PSID, allowing for a mass at 0 and log-normal distributions either side of 0. The distribution allows for correlation between the initial wealth endowment and the fixed component of individual wages (α_i).[28]

Estimation with first moments. The parameter estimates for the model are shown in column 1 of Table 1 and the fit of the model is displayed in Figure 1. The estimated model fits the data on mean hours and mean consumption well. The fit is good enough to easily pass a test of overidentification restrictions (p-value $= 0.45$).

Nonhuman wealth is crucial. Having access to financial wealth, and particularly the ability to borrow at young ages, is the key ingredient of the model that allows it to match the data on life-cycle means. To illustrate this fact, Figure 1 also shows the fit of

[28]The estimated wealth distribution matches the cross-section distribution of net worth for households with experience levels 3–5. When the model is estimated with the borrowing constraint set to zero, all households are given initial wealth greater than or equal to 0.

the model when the borrowing limit is set to zero and the parameters are reestimated (shown in column 2 of Table 1). Without borrowing, the model is unable to generate an upward sloping profile for mean hours and vastly overstates the growth in mean consumption. The test of overidentifying restrictions is strongly rejected (p-value $= 0.0$). The reason is that young households face relatively low wages and hold relatively small amounts of wealth. Borrowing allows the young to substitute their labor supply to later periods in life when wages will be higher, yet maintain only moderate consumption growth by borrowing from their future earnings.

5. Second moments: A challenge for the benchmark model

Although the model is able to successfully reproduce the evolution of mean hours and mean consumption over the life cycle, in this section I argue that the model fails with respect to the second moments of hours and consumption, regardless of whether these are included as targets for estimation.

Benchmark model fit on second moments. The fit of the benchmark model for the variances of hours and consumption are displayed in Figure 2.[29] The model slightly overstates the rise in consumption inequality over the life cycle, although the model's predictions lie within the 95% confidence bands. However, the model fails to generate the sharp decline in the variance of hours over the first 15 years. The reason that the model *can* generate a small reduction in the hours inequality over experience levels 3–10 is due to heterogeneity in initial wealth. This heterogeneity generates cross-sectional differences in the relative strengths of the income and substitution effect of wage shocks, which in turn generates heterogeneity in hours worked. As households accumulate financial wealth for life-cycle and precautionary reasons, the effects of these initial wealth differences dissipate. However, this effect is far too small to account for the decline in the data.[30]

Estimation including second moments as targets. Since the model is overidentified with data on life-cycle means, it is possible that by reestimating the model to target all four moments simultaneously, the fit could be improved. When doing this, I include the variance of measurement error in earnings, hours, and consumption as additional parameters to be estimated. The resulting parameter estimates are shown in column 3 of Table 1 and the fit of the model is displayed in Figure 1 (first moments) and Figure 2 (second moments). The model fit and parameter estimates are essentially unchanged from the estimation that used only first moments. The test of overidentifying restrictions

[29] In Figure 2, the *level* of inequality is the same in the data and the model for both consumption and hours. This is because classical measurement error in consumption and hours is assumed to be responsible for any positive difference between the cross-sectional variance in the data compared with the model. This procedure implicitly generates an estimate of the amount of measurement error in hours and consumption. When I extend the estimation exercise to include the second moments as targets, I explicitly include the variances of measurement error as parameters to be estimated.

[30] There is also a second contributing factor: the small decline in wage inequality at young ages. This is due to the fact that the estimated variance of transitory wage shocks is falling very early on in the life cycle. See Figure 16 in Appendix E.

is now overwhelmingly rejected (p-value $= 0.0$). The next section evaluates the source of this failure.[31]

6. Implications from the first-order condition for labor

In this section, I use the intratemporal first-order condition to show that the failure of the benchmark model to generate a significant decline in hours inequality over the first half of the working life is a general shortcoming of this class of models. I then infer that to improve the model in this dimension, a modification that generates a labor wedge with declining cross-sectional variance is required. For ease of exposition, I consider the preference specification of Section 2. In Appendix H, I show that the argument generalizes to other reasonable preferences, including those that allow for nonseparability between consumption and hours.

Intratemporal first-order condition. The first-order condition for the choice of hours in the benchmark model is

$$w_{it}c_{it}^{-\gamma} = \varphi_i H_{it}^{\sigma}. \tag{1}$$

By taking logs and cross-sectional variances, it is possible to arrive at an expression that relates the cross-sectional variance of hours worked to moments of the joint distribution of wages, consumption, and preference heterogeneity:

$$\begin{aligned}\sigma^2 V(\log H_{it}) = {} & V(\log w_{it}) + \gamma^2 V(\log c_{it}) + V(\varphi_i) \\ & - 2\gamma \operatorname{COV}(\log c_{it}, \log w_{it}) + 2\gamma \operatorname{COV}(\log c_{it}, \log \varphi_i).\end{aligned} \tag{2}$$

Taking first differences gives the useful relationship

$$\begin{aligned}\sigma^2 \Delta V(\log H_{it}) = {} & \Delta V(\log w_{it}) + \gamma^2 \Delta V(\log c_{it}) \\ & - 2\gamma\Delta \operatorname{COV}(\log c_{it}, \log w_{it}) + 2\gamma \operatorname{COV}(\Delta \log c_{it}, \log \varphi_i).\end{aligned} \tag{3}$$

A declining variance of hours at young ages corresponds to a negative value for the left hand side of (3). However, in the data, the variance of wages is sharply increasing over the life cycle while the variance of consumption is either flat or increasing. Hence the first two terms on the right hand side of (3) are positive.

The covariance between consumption and wages does indeed trend slightly upward over the life cycle, so it is possible that a negative component could arise from this term. However, simple back-of-the-envelope calculations confirm that the increase in $\operatorname{COV}(c_{it}, w_{it})$ in the data is not nearly large enough to generate the required decline in the variance of log hours. Moreover, to the extent that $\Delta V(\log c_{it}) > 0$, any attempt to

[31] Note that the convexity of the profile for the variance of log hours is far smaller in the model than the data. This is important since the second derivative of an age profile is always identified, regardless of which identification scheme for dealing with year/cohort effects one adopts (see Yang et al. (2008)). Hence the finding that the model is not consistent with the life-cycle profile of the variance of hours cannot be due to the choice of controlling for year effects, cohort effects, or any other possible choices for identification.

amplify the impact of $\mathrm{COV}(c_{it}, w_{it})$ by increasing γ is offset by a larger opposing contribution from the variance of log consumption, for values of γ above 2.

Finally, when borrowing is allowed (which was argued in Section 4 to be the necessary ingredient to match the life-cycle properties of mean hours), the Euler equation implies that consumption growth should not co-vary significantly with fixed individual characteristics. Hence the final term in (3) is too close to zero to have any meaningful impact on the variance of hours. Moreover, even if this term were larger, its sign is likely to be positive: agents with a high disutility of working are likely to be closer to their borrowing constraints and hence will have consumption that more closely tracks wages, which slopes upward.

This argument clarifies that the difficulty for the model is in *simultaneously* accounting for inequality in wages, hours, and consumption. It is the joint restrictions implied by this first-order condition that are inconsistent with the life-cycle patterns of inequality in the data. It is important to note that the exact specification of the stochastic process for wages does not affect this conclusion: *any* wage process that is consistent with the evolution of the cross-sectional variance of wages over the life cycle will suffer from this difficulty. Storesletten, Telmer, and Yaron (2001) showed that a similar argument holds when markets are complete.

An age-varying wedge? One mechanical way to generate a decline in $V(\log H_{it})$ would be to allow for age variation in the cross-sectional variance of preference heterogeneity. That is, if one were to allow for individual-specific shocks to the disutility of working, φ_{it}, it would be possible to recover an implied age path for the variance of these shocks that could rationalize the data on hours inequality.[32] One problem with such an approach is that it is hard to think of an underlying structural motivation for preference shocks whose variance declines with age. In Section 7, I argue empirically that the declining variance of hours is not driven by observable characteristics, so one could not appeal to marriage, fertility, or home ownership as the source of these differences.

Hence it is clear what type of modification is required so as to generate declining inequality in annual hours: a change that introduces a wedge in the first-order condition for labor whose cross-sectional variance decreases with age. In the sections that follow, I argue that age-varying labor market frictions, as implied by observed age differences in job destruction rates, can manifest themselves as a wedge with exactly this property.

7. In search of an age-varying wedge: Unemployment

In this section, I delve deeper into the age profile of inequality in hours to show that it is a robust feature of the data. First, I argue that the declining variance of hours is not driven by observable characteristics. Next I show that the decline is present along both the intensive and extensive labor supply margins. I then move beyond the variance of logs as a measure of inequality and argue that all of the decline in inequality occurs in the bottom half of the hours distribution. Finally, I show that there are important age

[32] Badel and Huggett (2010) formalized this intuition in a setting with complete markets. They showed that with enough age variation in the structure of preference shocks, it is possible to account for the age patterns in inequality in wages, hours, and consumption.

patterns in weeks spent unemployed during the year that imply age variation in workers' ability to freely choose hours of work during the year. Together, these findings motivate the extensions to the benchmark model that I explore in the remaining sections.

Controlling for observable characteristics. The sharp decline in the variance of annual hours with age is not driven by the effects of fixed or time-varying observable characteristics. To make this point, I use data from the Current Population Survey (CPS) and revert to using age (rather than potential labor market experience) as the measure of the life cycle.[33] Figure 3(a) shows the age profile for the variance of log annual hours in the full CPS sample together with various subsamples. The figure shows that there is a substantial decline in inequality up to the mid-30's *within* these subgroups: the decline occurs among married men, single men, those with children, those without children, those in the labor force, and those who own their own house.

To make this point more formally, Figure 3(b) shows the results of a standard between-group/within-group decomposition, where groups are defined based on a rich set of observable characteristics.[34] While some of the decline in the first few years is due to differences in observables, the vast majority of the decline comes from the within-group variance: cross-sectional variation in hours worked among individuals with similar observable characteristics.

Moreover, the decline in the within-group variance of hours is not due to compositional effects. In principle, the within-group variance of hours may fall because individuals change in predictable ways as they get older, with these changes implying lower hours inequality. For example, if the variance of hours is smaller among married men and the fraction of men who are married increases with age, then this will generate a fall in the overall variance of hours with age. To rule out this effect, I reconstruct the path of the within-group variance, removing the effect of compositional changes. The result is displayed in Figure 3(c).[35] This figure shows that only around one-third of the decline in the within-group variance of hours can be explained by this type of compositional bias.

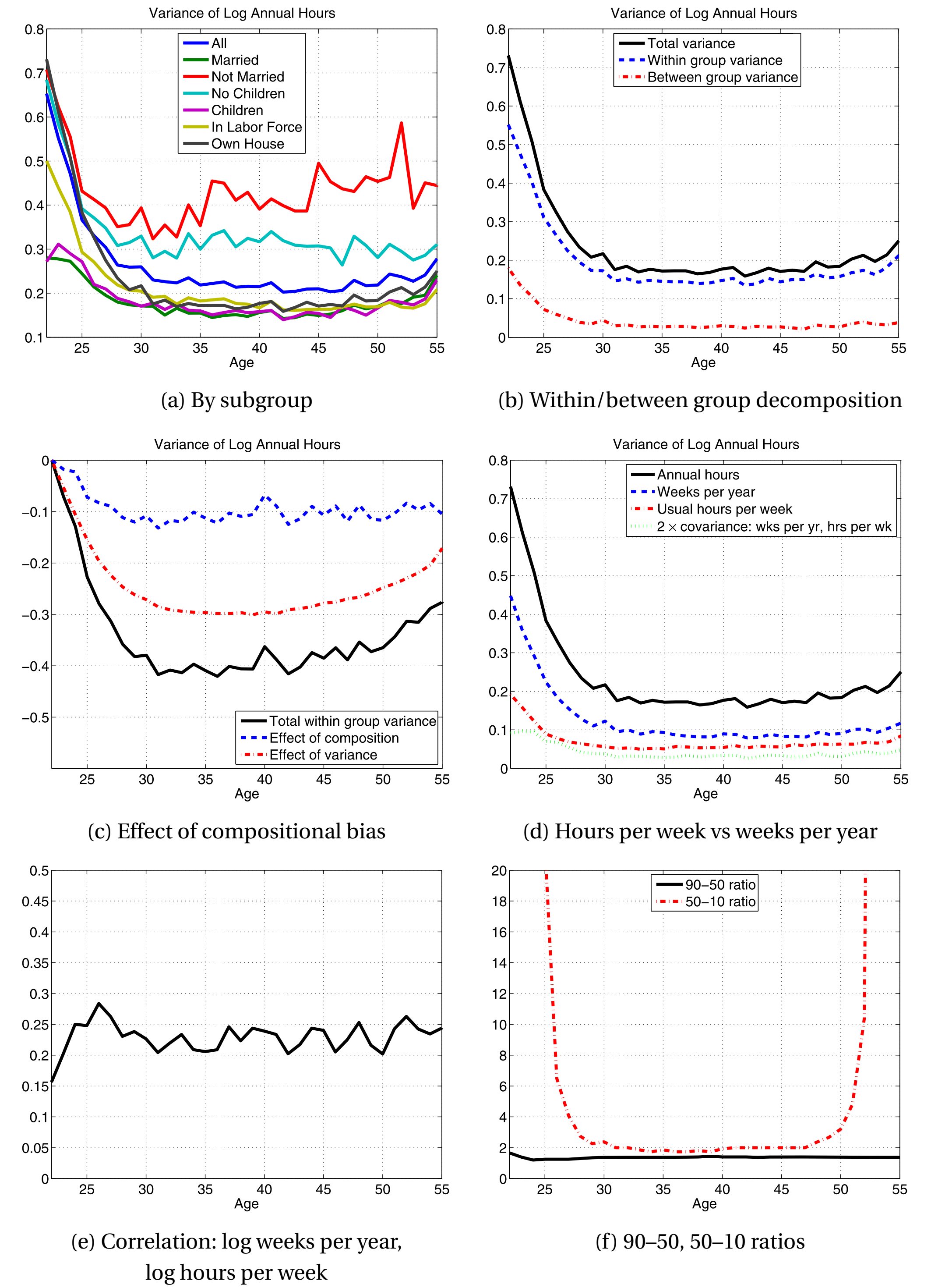
Extensive versus intensive margins. Recall that annual hours are measured as weeks worked per year multiplied by usual hours worked per week. Hence it is possible to decompose the variance of log annual hours as

$$\begin{aligned} V(\log H_{it}) = {} & V(\log[\text{usual hours per week}]) + V(\log[\text{weeks per year}]) \\ & + 2\,\mathrm{COV}(\log[\text{usual hours per week}], \log[\text{weeks per year}]). \end{aligned} \tag{4}$$

[33] The motivation for using the CPS is its substantially larger sample sizes compared with the PSID and CEX. This allows me to explore changes in the distribution of hours over the life cycle in more detail. The use of age as the life-cycle variable is for the sake of clarity only. In this section, the only selection criterion that I impose is to focus attention on males, and I document patterns in raw rather than residual inequality. Details on the CPS and the construction of the relevant variables can be found in Appendix A.

[34] The decomposition is $V(\log H_{it}) = E_X[V(\log H_{it} \mid X)] + V_X[E(\log H_{it} \mid X)]$. The first term reflects the "within-group" variance: cross-sectional variation in hours that exists among individuals with the same X characteristics. The second term reflects the "between-group" variance: cross-sectional variation due to differences in X characteristics across the population. A group (X) is defined as a cell based on employment, labor force status, education, marital status, number of children, home ownership, and race.

[35] See Appendix D for details on the removal of compositional bias from a within-group variance.



(a) By subgroup

(b) Within/between group decomposition

(c) Effect of compositional bias

(d) Hours per week vs weeks per year

(e) Correlation: log weeks per year, log hours per week

(f) 90–50, 50–10 ratios

Figure 3. Features of hours distribution over life cycle in CPS. See the text for a description of the plots.

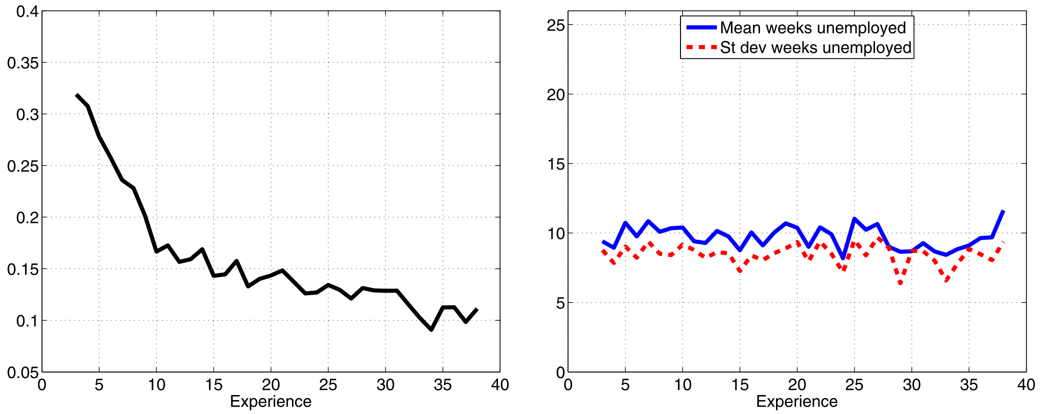
Figure 3(d) shows the results of this decomposition. Two interesting facts emerge. First, both usual hours per week and total weeks per year contribute to the fall in the variance of annual hours. The relative contribution is roughly 25% from hours per week, 65% from weeks per year, and the remaining 10% from the covariance term. Hence both the extensive and intensive margins are important for understanding the decline in inequality. Second, the cross-sectional covariance between usual hours per week and weeks per year is positive and varies less with age than either of the two variance terms. This implies that individuals who work fewer weeks of the year also work fewer hours in the weeks that they are working. The implied correlation between usual weekly hours and weeks per year, shown in Figure 3(e), is flat with age and is around 0.25.

Beyond variance of logs. Until now, I have only considered the variance of logs as a measure of inequality. Figure 3(f) displays two alternative measures: the 90–50 ratio and the 50–10 ratio. A striking feature immediately emerges: the decline in hours inequality over the life cycle takes place exclusively in the bottom half of the distribution of hours. Put differently, inequality is high for young workers because there is a group of young workers who work far fewer hours than the median, not because there are many young individuals who work particularly long hours.

Comparison with the literature. Badel and Huggett (2010) and Erosa, Fuster, and Kambourov (2011) also documented patterns in the distribution of hours worked over the life cycle. The evidence in both papers is consistent with that presented here. Badel and Huggett (2010) (Figure A.1) targeted an age profile for the variance of hours worked that features a smaller decline at young ages. This is due to two reasons. The first is that they restricted their sample to households that work at least 23 hours per week and 13 weeks per year. This reduces the decline, consistent with the evidence in Figure 3(f). The second reason is that they average across 5-year age groups, which has the effect of flattening any trend in the data. Their model can generate only a flat or increasing profile for the variance of hours. The same is true for Erosa, Fuster, and Kambourov (2011), who documented a decline (Figure 2) at young ages similar to the one that I do, but targeted the age profile of hours only at ages 26 and over, over which range the coefficient of variation is flat and then starts to rise (Figure 10) as retirement nears.

Unemployment by age. The hypothesis that I pursue as an explanation for the decline in inequality at the bottom of the hours distribution is age differences in the incidence of unemployment. Since the aim is to incorporate the evidence on unemployment into the structural estimation, I revert to the PSID and focus on the same sample that is used in estimation. Since 1970, the PSID has collected data on the number of weeks spent unemployed in each calender year.[36] The main features of the data are displayed in Figure 4. First, Figure 4(a) shows that the incidence of unemployment decreases sharply with age: the fraction of the sample experiencing at least 1 week of unemployment during the calender year decreases by 20 percentage points, from around 25% to 5%. Second, Figure 4(b) shows that, conditional on experiencing an unemployment spell during the year, the distribution of weeks unemployed is remarkably constant with age. On av-

[36]See Appendix A for further details on the construction of the weeks unemployed variable.



(a) Fraction with at least 1 week unemployed

(b) Distribution of weeks unemployed, conditional on at least 1 week

Figure 4. Weeks unemployed during the year from PSID.

erage, 10 weeks are spent in unemployment with a standard deviation of 8.6 weeks.[37] In the remainder of the paper I argue that if unemployment spells are viewed as periods when workers may like to work but are constrained from doing so, then involuntary unemployment can provide the missing ingredient that is needed to match the data on hours inequality, both in terms of the extensive and intensive margins.

8. Modelling unemployment in an annual model

In this section, I introduce a role for unemployment shocks into the model of frictionless hours choices outlined in Section 2. Since the vast majority of unemployment spells are much shorter than 1 year, a natural way to think about unemployment in an annual model is as reductions in the available time endowment. In other words, unemployment spells are periods within the year when the individual is prohibited from working. My approach is to treat unemployment shocks as exogenous reductions in the annual time endowment, where the stochastic process that drives these reductions is calibrated based on the evidence presented in Section 7. The resulting model is thus a hybrid between one that treats labor supply as a choice (the frictionless intensive margin) and a constraint (the frictional extensive margin).[38]

Shocks to the time endowment. Let $\overline{h}_i \in [0, 1]$ be an individual's time endowment for employment in a given year. Thus $1 - \overline{h}_i$ is the fraction of the year spent unemployed. Each individual can choose to work a fraction $h_i \in [0, 1]$ of his available time endowment. Hence annual hours worked are $H_i = \overline{h}_i h_i$ and annual earnings are $w_i h_i \overline{h}_i$.

[37] These facts are consistent with a job separation rate that declines with age but a job finding rate that is constant with age. A number of authors have found evidence for this in other data sets. For example, using data from the Survey of Income and Program Participation, Menzio, Telyuokva, and Visschers (2010) documented that the unemployment–employment transition rate is flat with age and that the employment–unemployment transition rate declines sharply with age.

[38] For an alternative approach to modelling intensive and extensive labor supply margins simultaneously, see Rogerson and Wallenius (2009) and Prescott, Rogerson, and Wallenius (2009).

Preferences over $(h_i, \overline{h}_i)$. A natural starting point would be to adopt the preference specification from the benchmark model, where agents have preferences defined over total annual hours, H_i. In that case, the disutility of work is given by

$$v(h, \overline{h}) = \frac{(\overline{h}h)^{1+\sigma}}{1+\sigma}. \tag{5}$$

However, modelling preferences in this way has two unappealing features. First, these preferences imply that agents are indifferent about when in the year they work, and how their annual work hours are spread across periods of employment and unemployment. For example, a worker with these preferences would be indifferent between working 42 hours a week for the entire year (42×52 weeks $= 2184$ hours) versus working 24 hours a day for only 3 months of the year (24×91 days $= 2184$ hours). This is clearly counterintuitive: people do indeed care about how their work hours are spread throughout a year.

Second, these preferences imply that exogenous reductions in the time endowment will have little or no effect on the distribution of total hours worked. To see this, recall that the model is calibrated to match mean annual hours worked across individuals, which is approximately 40% of the available total time endowment. This means that there are virtually no agents whose hours choices are constrained by their time endowment. Since no agents are constrained, tightening this constraint does not have an effect on allocations.

The economics of this implication is that with preferences defined over *total hours*, a reduction in the total time endowment effectively allows individuals to choose in which of their annual hours they will be prevented from working. Since only 40% of available hours are spent working, there is ample scope to locate these periods of unemployment at times when they would not want to work anyway.

A richer preference specification. To generate more realistic implications of reductions in the time endowment, I consider preferences that display nonseparability between time spent unemployed and hours worked during employed periods. I assume the specification

$$v(h, \overline{h}) = \frac{(\overline{h}^{\chi} h)^{1+\sigma}}{1+\sigma}. \tag{6}$$

Special cases. To understand the implications of these preference, it is easiest to consider four special cases:

(i) When $\chi = 1$, preferences are identical to those in (5). Individuals care only about total hours worked during the year.

(ii) When $\chi = \frac{1}{1+\sigma}$, preferences are linear in the time endowment, that is, $v(h, \overline{h}) = \overline{h}\frac{h^{1+\sigma}}{1+\sigma}$. These are the preferences that would arise if one modelled the disutility of working in each employed week as being separable across weeks within the year, with the restriction that hours worked in each employed week are the same.

(iii) When $\chi = 0$, individuals do not derive disutility from additional periods of employment, but only from the intensity of hours worked while employed. Hence individuals only care about hours worked per employed week. One could obtain this preference specification by assuming either that no leisure is enjoyed during unemployment spells or that leisure during unemployment spells is separable from hours worked during employment.

(iv) When $\chi < 0$, periods of unemployment generate additional disutility of working, over and above the additional hours of work required to keep total annual hours constant. This implies that weeks spent in unemployment increase the disutility of each hour of work even when employed. There are a number of ways that one could rationalize this sort of behavior. For example, in a weekly model of employment, habits in labor supply would lead individuals who spend more weeks unemployed to work fewer hours in the weeks that they are working. Similarly, if individuals learn about job termination in advance of stopping work, then it is reasonable to think that the incentives for them to work long hours in the weeks leading up to the end of the employment spell are reduced.

These preferences are intended to capture, in a reduced form way, some important features of labor supply that are missing from (5). First, there are both *choice* and *constraint* elements to labor supply within the year. The utility function (6) can be thought of as an attempt to bridge the gap between neoclassical models of labor supply that assume that workers are free to choose to work as many hours as they wish at a given a wage, and frictional models where employment offers are the result of random outcomes of search effort.[39] Some of the observed cross-sectional variation in hours is due to different choices of hours worked across individuals, while part is due to different opportunities to work. Second, this model takes seriously the idea that people care about how their work hours are spread across periods of employment and unemployment.

Impact on $V(\log H)$. With unemployment modelled in this way, the first-order condition for the choice of hours becomes

$$w_{it} c_{it}^{-\gamma} = \overline{h}_{it}^{(\chi-1)(1+\sigma)} \varphi_i H_{it}^{\sigma}. \tag{7}$$

Thus $\overline{h}_i^{(\chi-1)(1+\sigma)}$ acts as an unemployment wedge. Since the incidence of unemployment decreases strongly with age, the cross-sectional variance of this wedge also decreases with age. It is possible to derive an analogous expression to (3) for the change in the cross-sectional variance of total hours worked,

$$\begin{aligned}\sigma^2 \Delta V(\log H_{it}) = {}& \Delta V(\log w_{it}) + \gamma^2 \Delta V(\log c_{it}) \\ & - 2\gamma \Delta \operatorname{COV}(\log c_{it}, \log w_{it}) + 2\gamma \operatorname{COV}(\Delta \log c_{it}, \varphi_i) \\ & + (\chi - 1)^2 (1+\sigma)^2 \Delta V(\log \overline{h}_{it}) \\ & + 2\gamma(\chi - 1)(1+\sigma) \Delta \operatorname{COV}(\log \overline{h}_{it}, \log c_{it}),\end{aligned} \tag{8}$$

[39]Note that this model of labor supply has the flavor of a search model in that for some part of the year, individuals may not have an offer to work, and when they do, they are free to either work zero or a positive number of hours. What the model lacks relative to traditional search models is an option value of waiting—there are no dynamics in labor supply other than through the effects of consumption and asset accumulation.

where the last two terms reflect the additional impact of shocks to the time endowment. With loose borrowing constraints, the final term is close to zero since shocks to the time endowment are transitory and are thus easily smoothed.

The important term is $(\chi - 1)^2(1 + \sigma)^2 \Delta V(\log \overline{h}_{it})$. Its impact on the variance of total hours comprises two parts. First there is the direct effect from changes in annual employment, $\Delta V(\log \overline{h}_{it})$. The second effect stems from intertemporal nonseparability in labor supply. When $\chi = 1$, these two effects exactly offset and there is no impact on the variance of hours worked. When $\chi > 0$, there are only partially offsetting changes in intensive hours so that the total effect is positive but smaller than the direct effect. When $\chi < 0$, there may be amplification: a spell of unemployment reduces hours worked even in employed periods, thus increasing cross-sectional variation in annual hours by more than the direct effect from cross-sectional differences in employment opportunities. When $\chi = 0$, this offsetting is just enough so that only the direct effect operates. Hence to generate additional inequality in annual hours over and above inequality from unemployment, it is necessary to have $\chi < 0$.

Allowing for "lazy" types. Finally note that if one allows for correlation between the disutility of work and unemployment shocks, there may be an additional effect. The following extra term is introduced to (8):

$$+2(\chi - 1)(1 + \sigma)\Delta \operatorname{COV}(\log \overline{h}_{it}, \log \varphi_i).$$

If this covariance is negative, then the intuition is that some people are lazy types. These individuals both work less in periods when employed and are more likely to experience unemployment spells. Since unemployment becomes less likely as individuals age, the contribution of this component declines with age. Hence this term has a negative contribution to the overall change in the variance of annual hours whenever $\chi \leq 1$: with lazy types, amplification is possible even if $\chi \geq 0$.

9. Estimation with unemployment shocks

I now embed the model of labor supply with unemployment shocks from Section 8 into the incomplete markets life-cycle model from Section 2. The strategy is to calibrate an exogenous process for the time endowment from the data on weeks spent unemployed during the year and then repeat the estimation from Section 4 using the richer model. As in Section 4, the targeted moments are the means and variances of log hours and log consumption by potential experience.

Calibration of unemployment shocks. I treat unemployment as an exogenous shock that reduces the time endowment in a given year. The evidence presented in Section 7 shows that all the age variation in weeks spent unemployed is due to the incidence of an unemployment spell, rather than in the number of weeks spent unemployed, conditional on experiencing an unemployment spell during the year. Thus I model unemployment shocks as an age-varying probability of suffering a reduction to the time endowment ($\overline{h}_{it} < 1$), with probabilities taken directly from Figure 4. Conditional on receiving such a shock, $\overline{h}_{it}$ takes on one of two possible values with equal probability. The two values are set so that the expected reduction in the timing endowment is 20% with a standard deviation of 17%, consistent with the evidence in Figure 4.

I assume that $\overline{h}_{it}$ is distributed independently over time. This assumption may at first appear too strong. In the data, the average autocorrelation of $\log \overline{h}_{it}$ is around 0.30. However, the majority of this autocorrelation is driven by fixed unobserved heterogeneity. The corresponding autocorrelation, controlling for individual-specific fixed effects, is 0.07 and is not significantly different from zero. Hence the richer model that allows for correlation between unemployment shocks and the disutility of working implicitly captures this autocorrelation.

Estimating χ. I start by assuming that $\overline{h}_{it}$ is identically distributed across the population and is independent of the disutility of work φ_i. The discussion in Section 8 made it clear that to match the data on the variance of hours, a negative value of χ is needed. The parameter estimates in column 1 of Table 2 show that this is indeed the case. The estimated value for χ is -0.36, suggesting that there is a small degree of intertemporal nonseparability in preferences over hours worked within the year. Note that χ is only slightly negative, so only a small amount of feedback from unemployment to the intensive hours choice is required to match the data. The negative value implies a positive cross-sectional correlation between the work time endowment and hours worked during the time when work is available. This is consistent with the evidence presented in Section 7. One way to quantify the size of this feedback is to ask by how much a worker would have to reduce his hours in employed periods to be indifferent about suffering an unemployment spell during the year, holding consumption constant. With the assumed preferences, it can be shown that this amount is given by $\overline{h}\exp(-\chi(1+\sigma)^{-1})$. At the estimated parameter values ($\sigma = 2.1$, $\chi = -0.36$), a 20% reduction in the time endowment generates a reduction in hours of 10.1%.

The fit of the model is displayed in Figure 5 (first moments) and Figure 6 (second moments). In contrast to the model that excluded unemployment shocks, the fit of the model is excellent. The test of overidentifying restrictions is easily passed, with a p-value of 0.28.

Estimating correlation between unemployment disutility of work. Some readers may find the idea of a negative value for χ unappealing. While there are a number of possible stories that could be used to micro-found a model with $\chi < 0$ (e.g., habits in labor supply, loss of motivation) I acknowledge that it may be more attractive to account for the data without appealing to this feature of the model. One way to achieve this is to allow for correlation between the disutility of working and the likelihood of experiencing an unemployment shock. This is the story of lazy types described in the previous section, and is also consistent with the observed autocorrelation of weeks unemployed being accounted for by individual fixed effects.

To estimate this version of the model, I fix $\chi = 0$, and allow the probability of suffering an unemployment spell to be related to the disutility of working, φ_i, using the formula

$$\Pr(\overline{h} < 1 \mid \varphi_i, t) = \left(1 + \rho \frac{\varphi_i}{E[\varphi]}\right)\pi_t,$$

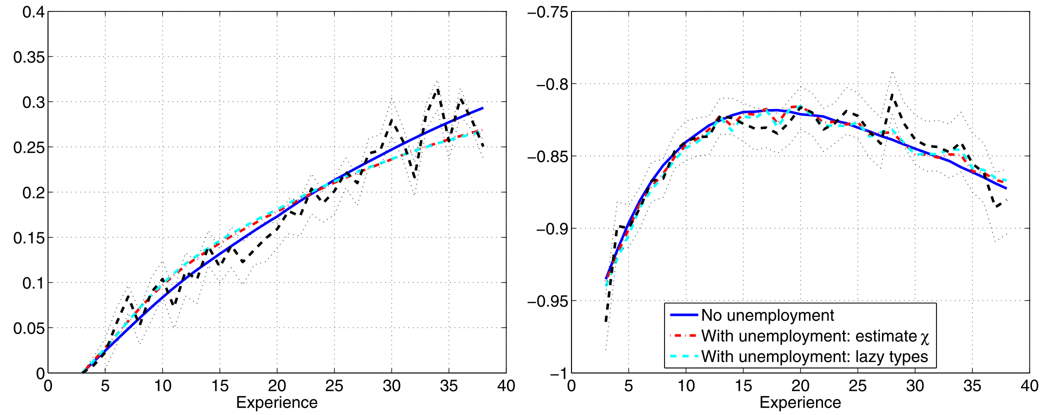
TABLE 2. Parameter estimates with shocks to time endowment.[a]

	(1)	(2)	(3)	(4)
Estimated parameters				
γ	2.33	2.47	2.24	2.49
	(1.86, 3.48)	(1.82, 3.68)	(2.06, 2.92)	(2.03, 2.96)
σ	2.10	2.26	2.41	3.05
	(1.49, 3.69)	(1.56, 4.27)	(1.72, 3.95)	(1.75, 4.44)
CV[φ]	0.21	0.23	0.80	0.77
	(0.00, 0.52)	(0.08, 0.82)	(0.71, 0.94)	(0.71, 0.93)
$\underline{a}$	−0.22	−0.22	−0.17	−0.19
	(−0.44, −0.12)	(−0.45, −0.11)	(−0.41, −0.04)	(−0.39, −0.07)
$\sigma^2_{\epsilon,y}$	0.02	0.01	0.10	0.06
	(0.00, 0.20)	(0.00, 0.07)	(0.04, 62.32)	(0.04, 15.48)
$\sigma^2_{\epsilon,h}$	0.02	0.02	0.00	0.00
	(0.00, 0.04)	(0.00, 0.04)	(0.00, 0.01)	(0.00, 0.00)
$\sigma^2_{\epsilon,c}$	0.08	0.08	0.06	0.06
	(0.06, 0.09)	(0.06, 0.11)	(0.05, 0.07)	(0.05, 0.07)
χ	−0.36		−0.08	
	(−0.52, −0.23)		(−0.22, 0.44)	
$\rho_{\varphi,\overline{h}}$		0.74		0.24
		(−0.02, 1.09)		(−0.26, 0.58)
Test of overidentifying restrictions				
p-value	0.28	0.26	0.00	0.00
Calibrated parameters				
β	0.989	0.989	0.989	0.987
$E\varphi$	66.4	88.4	115.1	305.0
Targeted moments				
$E[\log c]$	X	X	X	X
$E[\log h]$	X	X	X	X
$V[\log c]$	X	X	X	X
$V[\log h]$	X	X	X	X
COV[$\log w, \log h$]			X	X
COV[$\log c, \log h$]			X	X
COV[$\log w, \log c$]			X	X

[a] 95% bootstrap confidence intervals are given in parentheses. The borrowing limit $\underline{a}$ is expressed as a multiple of the average annualized hourly wage. $\sigma^2_{\epsilon,y}$ and $\sigma^2_{\epsilon,h}$ are estimates of classical measurement error in earnings and hours, respectively; when not estimated, they are set at 0.01. $\rho_{\varphi,\overline{h}}$ reflects the correlation between unemployment shocks and disutility of hours worked, as described in the text.

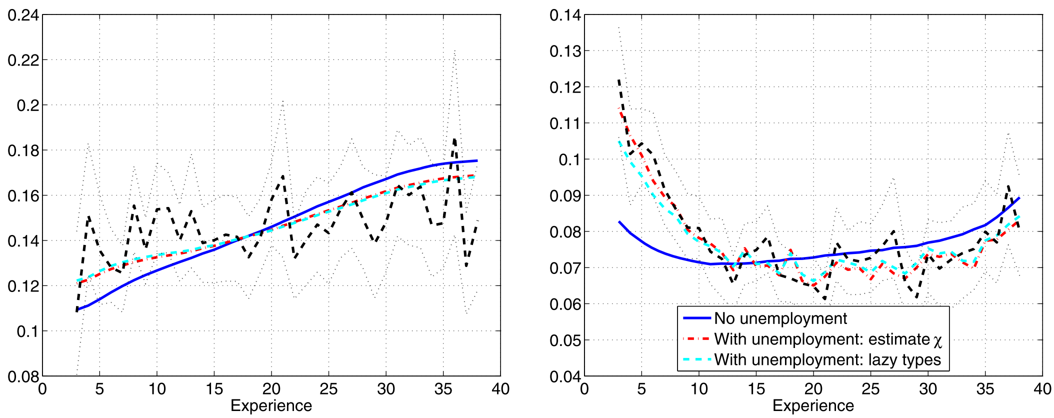
where π_t is the unconditional probability of an unemployment shock at experience t. When $\rho = 0$, the model collapses to the one in the previous section. When $\rho > 0$, individuals with above average values of φ_i are more likely to be unemployed. I estimate ρ along with the other parameters of the model.

The results of the estimation are displayed in Figures 5 and 6. The estimated value of ρ is 0.74. This model also does an excellent job of matching the data on consumption and hours. Again the test of overidentifying restrictions is easily passed (p-value $=$ 0.26).



(a) Mean log consumption (b) Mean log hours

FIGURE 5. Model fit for means (with unemployment). The thick dashed line is the data; the thin dotted line is the 95% confidence interval for the data; the solid line is the fit of the model without unemployment shocks; the dash-dot line is the fit of the model with unemployment shocks, without correlation between unemployment and disutility of labor; and the dashed line is the fit of the model with $\chi = 0$ and estimated correlation between unemployment and disutility of labor. All models target means and variances jointly.



(a) Variance log consumption (b) Variance log hours

FIGURE 6. Model fit for variances (with unemployment). The thick dashed line is the data; the thin dotted line is the 95% confidence interval for the data; the solid line is the fit of the model without unemployment shocks; the dash-dot line is the fit of the model with unemployment shocks, without correlation between unemployment and disutility of labor; and the dashed line is the fit of the model with $\chi = 0$ and estimated correlation between unemployment and disutility of labor. All models target means and variances jointly.

10. Further evaluation of the model

The focus of this paper has been on developing a theory that can simultaneously account for the life-cycle means and variances of consumption and labor supply. To this end, the model presented in Section 9 is largely successful.

However, as is true with all models, the model with unemployment shocks cannot be consistent with all aspects of the data. So in this section, I evaluate the fit of the model along some additional dimensions. There are two natural starting points. First, I consider the remaining second moments: covariances between wages, hours, and consumption. Second, I examine the key elements of the evolution of the wealth distribution over the life cycle.

Fit of covariances. The fit of the model for the joint second moments of wages, hours, and consumption is shown in Figure 7.[40] The model is successful in generating the broad pattern of a negative and downward sloped covariance between wages and hours. The reason is that with the estimated intertemporal elasticity of substitution $\frac{1}{\gamma} < 1$, the income effect of hours responses to wage changes is negative. Agents face two types of wage shocks in the model. There is little or no income effect associated with transitory shocks. Hence hours co-vary positively with transitory shocks due to the substitution effect. Persistent shocks, however, do generate an income effect and hence co-vary negatively with wages. As households age, persistent shocks account for an increasing fraction of the overall cross-sectional variance of wages. This generates a declining covariance between log wages and log hours in the model.

The model is less successful at matching the covariances of wages and hours with consumption. The model predicts a counterfactually negative covariance between hours and consumption, and overstates the rise in the covariance between wages and consumption.

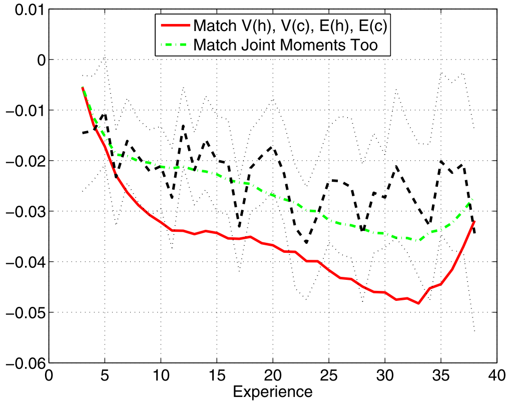
These shortcomings can be mostly overcome by reestimating the model to target the three covariances, in addition to the mean and variance of consumption and hours.[41] The parameter estimates are shown in columns 3 and 4 of Table 2, and the fit is displayed in Figure 7. Although the reestimated model fails the test of overidentifying restrictions (p-value $= 0.00$), the fit of the three covariances is substantially improved and the model is broadly consistent with all features of the joint distribution of wages, hours and consumption up to second moments.[42]

Fit of intensive and extensive margins. Figure 8 shows the fit of both models along the intensive and extensive margins, and the covariance between the two. These moments

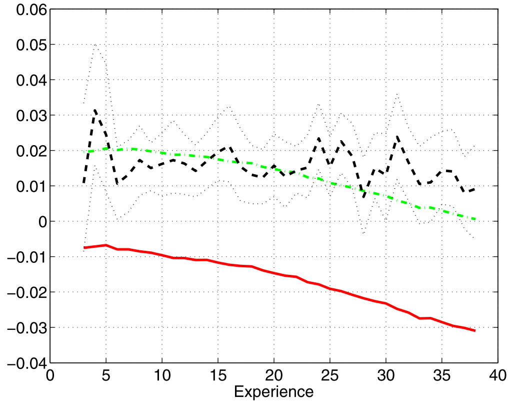
[40]The figures show the fit for the model that estimates χ. Along these dimensions, the fit of the lazy types model is extremely similar. Results are available from the author on request.

[41]Recall that the model is still overidentified when targeting the means and variances, and passes the test of overidentifying restrictions. Hence there is scope for an alternative parameter configuration to also match the three covariances. This turns out to be achieved in estimation by a very slightly worse fit for mean hours and the variance of log hours, but the change is so small it can barely be detected by the naked eye. Figures are available from the author on request. Statistically the fit is no worse along these dimensions, since it still passes the test of overidentification ignoring the three joint moments.

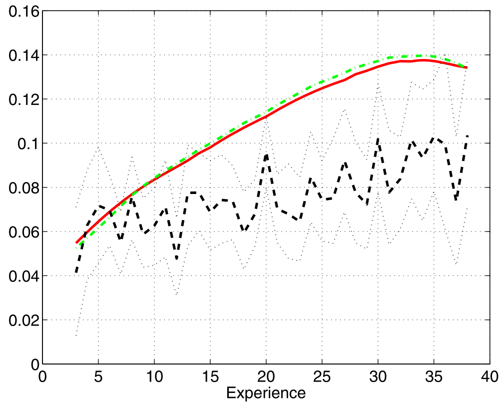
[42]Inspection of Figure 7(c) reveals why the model still fails the formal test of overidentifying restriction: the required increase in the covariance between wages and consumption that is required to match the other moments is too large in the model compared with the data.



(a) Covariance log wages, log hours



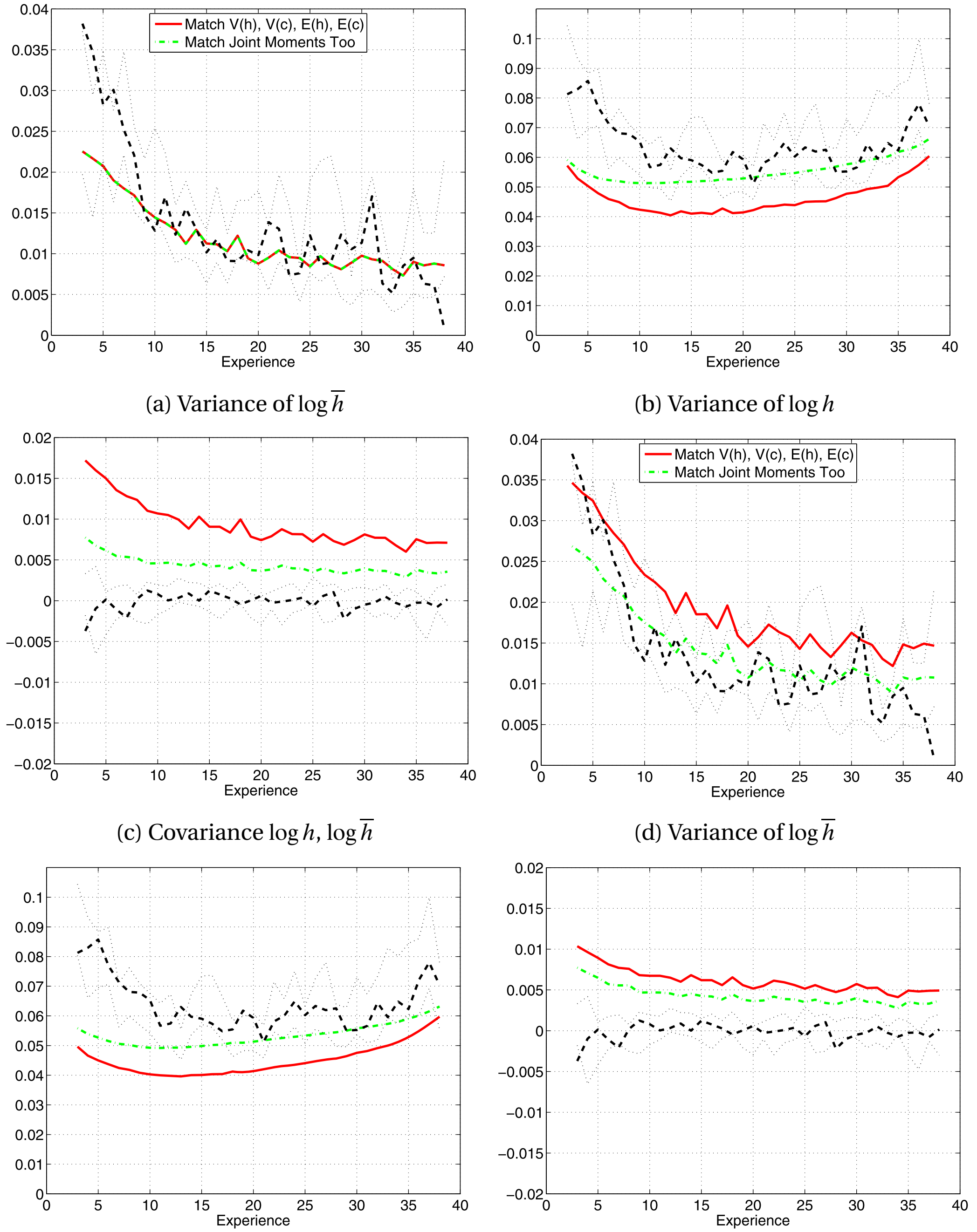
(b) Covariance log consumption, log hours



(c) Covariance log wages, log consumption

FIGURE 7. Model fit for covariances (with unemployment, estimating χ). The thick dashed line is the data; the thin dotted line is the 95% confidence interval for the data; the solid line is the fit of the model when matching means and variances only; and the dash-dot line is the fit of the model when matching means, variances, and covariances. All models estimate χ and do not allow for correlation between unemployment and disutility of labor.

are not targeted in estimation. The model does fairly well along all dimensions, particularly when the three joint moments described above are also targeted. A key difference between the model and the data is that the model features a declining, and slightly positive, covariance between the two margins, whereas in the PSID data, this covariance is zero and flat. However, this feature of the PSID data should be viewed with caution. In Section 7, I documented a positive and declining covariance between weeks worked and usual hours per week in the CPS, consistent with the model. The CPS is a far larger sample than the PSID and asks directly about these two margins. The PSID, on the other hand, has data on total hours and number of weeks unemployed. The profiles for usual hours per week in Figure 8 are constructed as the ratio of the two variables. Hence if there is any classical measurement error in the weeks unemployed variable in the PSID, it would bias the covariance downward, while the CPS measure would not be affected.



(a) Variance of $\log \overline{h}$

(b) Variance of $\log h$

(c) Covariance $\log h$, $\log \overline{h}$

(d) Variance of $\log \overline{h}$

(e) Variance of $\log h$

(f) Covariance $\log h$, $\log \overline{h}$

FIGURE 8. Model fit for intensive and extensive margins. The thick dashed line is the data; the thin dotted line is the 95% confidence interval for the data; the solid line is the fit of the model when matching means and variances only; and the dash-dot line is the fit of the model when matching means, variances, and covariances. Models in the top row estimate χ and do not allow for correlation between unemployment and disutility of labor. Models in the bottom row set $\chi = 0$ and estimate the correlation between unemployment and disutility of labor.

Fit of wealth distribution. The accumulation of nonhuman wealth for life-cycle and precautionary reasons is the key feature of the model that generates a dynamic aspect to labor supply choices and differing consumption choices for households of different ages. It is hence important that the model is consistent with the main features of the evolution of wealth over the life cycle. The fit of the model with regard to the wealth distribution is displayed in Figure 9. The key features are matched well: a declining fraction of the population with nonpositive wealth, a declining profile for the variance of log wealth (conditional on positive wealth), and a covariance between log wages and log wealth that is positive at all experience levels, increasing during the first half of the working life, and flat in the second half. The fit of the Gini coefficient is not as a good. The model generates a larger Gini than in the data in the first half of the working life, primarily due to the larger fraction of the population with zero or negative wealth. In the second half of the working life, the model understates the Gini, due to the model's inability to generate a fat upper tail of the wealth distribution. This is a well known shortcoming of this class of models with no entrepreneurial sector and no asset returns risk.

11. Concluding remarks

I have shown that an estimated life-cycle economy with incomplete consumption insurance and endogenous labor supply can successfully account for the observed joint distribution of consumption, wages, and hours in U.S. data. I argued that the intratemporal first-order condition for labor places very strong restrictions on how this distribution evolves over the life cycle and that these restrictions are inconsistent with the observed decline in hours inequality over the first half of the working life. These restrictions are so strong, that to fit the data, some modification to the standard model is needed. The modification that is required is one that generates an idiosyncratic labor wedge whose cross-sectional variance declines with age.

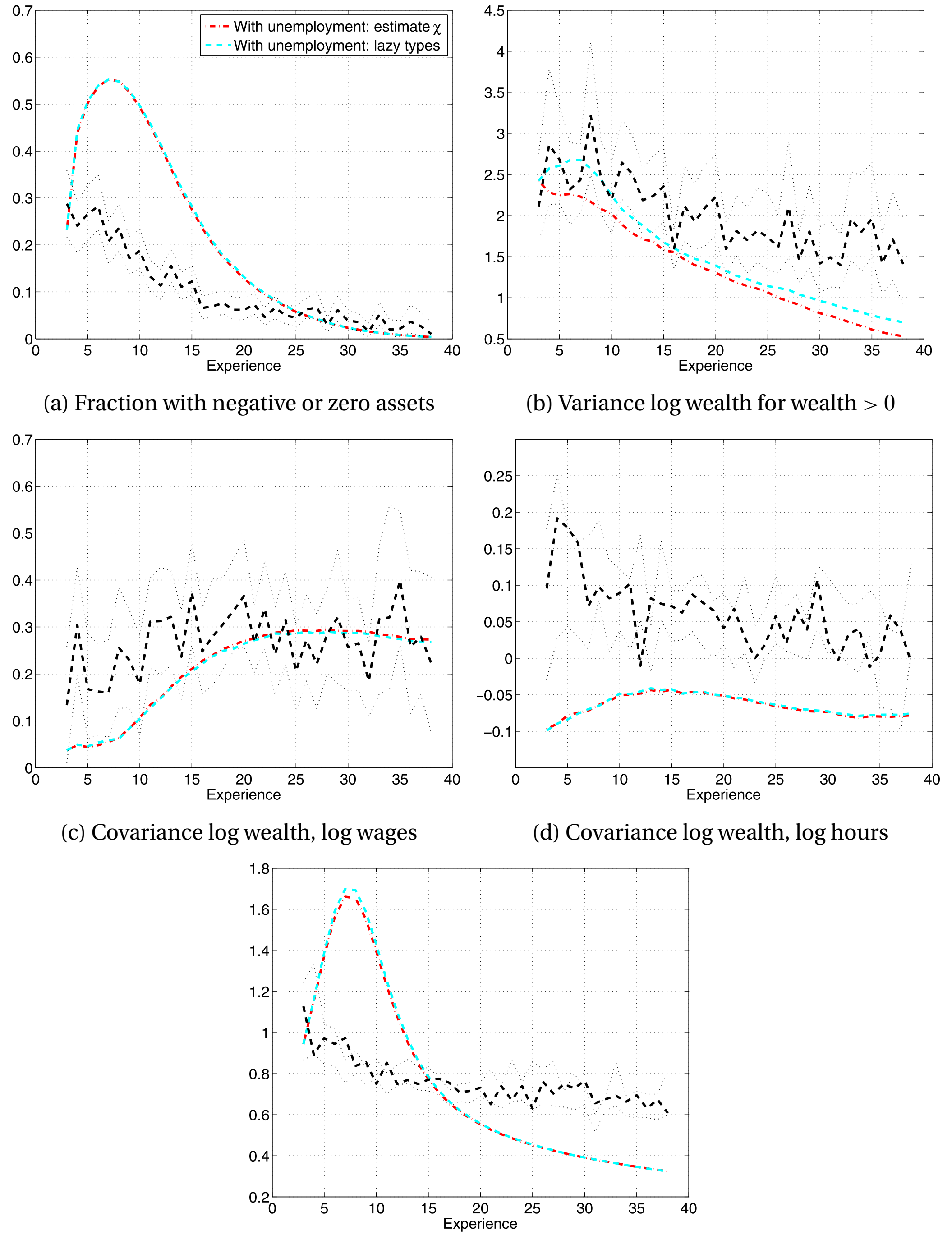
I modified the standard model to allow for the possibility of involuntary unemployment, which acts as shocks to the endowment of time available for work. For these shocks to generate a labor wedge with the desired cross-sectional properties, the model requires some degree of nonseparability in the disutility of hours worked at the extensive and intensive margins within a year: in the model, agents care about how their annual work hours are spread across periods of employment and unemployment. Thus the model generates inequality in annual hours worked that is partly due to differences across households in opportunities to work and partly due to differences in optimal labor supply decisions.

When the degree of unemployment risk is calibrated to be consistent with PSID data on unemployment spells, the estimated model matches the age profiles of the first and second moments of consumption, hours, and wages.

Appendix A: Data sources

A.1 *Panel study of income dynamics*

Data source. The PSID has been conducted annually from 1968–1997 and bi-annually since 1997. I use data from the 1970–2005 waves. Questions about labor income and



(a) Fraction with negative or zero assets

(b) Variance log wealth for wealth > 0

(c) Covariance log wealth, log wages

(d) Covariance log wealth, log hours

(e) Gini coefficient for wealth

FIGURE 9. Model fit for wealth distribution (with unemployment). The thick dashed line is the data; the thin dotted line is the 95% confidence interval for the data; the dash-dot line is the fit of the model with unemployment shocks, without correlation between unemployment and disutility of labor; and the dashed line is the fit of the model with $\chi = 0$ and estimated correlation between unemployment and disutility of labor. All models target means and variances jointly.

work hours refer to the previous calender year. Thus the 1970–1997 waves contain data on earnings in 1969–1996, and the 1999–2005 waves contain bi-annual data on earnings from 1998–2004. I do not use the 1968 or 1969 waves because data on weeks spent unemployed only became available in the 1970 wave.

The PSID data are released in two stages: an early release file with variables named ERxxxxx, and a final release file with variables named Vxxxx. The final release file contains data that have been subject to more stringent cleaning and checking processes, and contains a number of constructed variables (e.g., total annual labor income of the head and spouse). From 1994 on, the final release files have not been made available. Instead, clean variables for labor income, annual hours, and several other variables are available in the Family Income-Plus Files and Hours of Work and Wage Files. I restrict attention to the core Survey Research Center (SRC) subsample. This selection is imposed by retaining only those households with identifiers of 3000 or below.

Top-coding. From 1968 I deal with top-coded observations by assuming that the underlying distribution for each component of income is Pareto, and by forecasting the mean value for top-coded observations by extrapolating a Pareto density fitted to the non-top-coded upper end of the observed distribution.

Variable definitions. Annual earnings of the household head include all income from wages, salaries, commissions, bonuses, overtime, and the labor part of self-employment income. The PSID splits self-employment income into asset and labor components using a 50–50 rule. Annual hours of work are defined as the sum of annual hours worked on the main job and on extra jobs plus annual hours of overtime. It is computed by the PSID using information on usual hours worked per week and the number of actual weeks worked in the last year. Hourly wages are constructed as annual earnings divided by annual hours. Weeks unemployed are defined as the number of weeks of worked that were missed due to unemployment or temporary layoff.

Wealth data. The most reliable source for data on wealth inequality is the Survey of Consumer Finances (SCF). However, to avoid the use of a fourth data source and to maintain a sample that is consistent with the one used for data on consumption, hours, and wages, I use wealth data from a supplemental questionnaire in the PSID that was administered to a subsample of the panel in 1984, 1989, 1994, 1999, 2001, 2003, and 2005. Moreover, the use of PSID allows for the inclusion of evidence on the covariance of wealth with wages and hours over the life cycle, for which the model provides strong predictions. I define wealth as total net worth of a household. Although the model features only a single risk-free asset with which households can save, the use of total wealth, which includes elements such as housing and other nonliquid assets, is justified on two grounds. First, it is likely that almost all assets held by households can be liquidated within a year. Since the model period is annual, this is consistent with including seemingly nonliquid assets in the definition of wealth. Second, running down assets to smooth the effects of wage shocks in the model can be thought of in terms of borrowing against collateralizable assets in the real world.

Selection criteria. A household is retained in the sample in every year that it satisfies all of the selection criteria. Thus households may drop in and out of the sample. The following selection criteria are imposed: (i) retain only SRC sample; (ii) keep only heads of households; (iii) keep only males; (iv) drop observations with missing data on years of education; (v) drop 1968 and 1969 waves; (vi) keep only individuals aged between 20 and 60; (vii) keep only individuals with between 3 and 38 years of potential labor market experience; (viii) drop households with a second earner who earned at least half the amount earned by the male head—this is the single primary earner assumption discussed in the main text; (ix) keep only individuals who worked between 520 and 5200 hours during the calender year; (x) drop observations where the nominal wage is less than half the corresponding minimum wage for that year; (xi) drop 13 observations with nominal earnings greater than $1,000,000. The final sample contains 46,369 individual/year observations and 5469 distinct individuals. Table 3 shows the number of observations and individuals lost at each stage of the selection process.

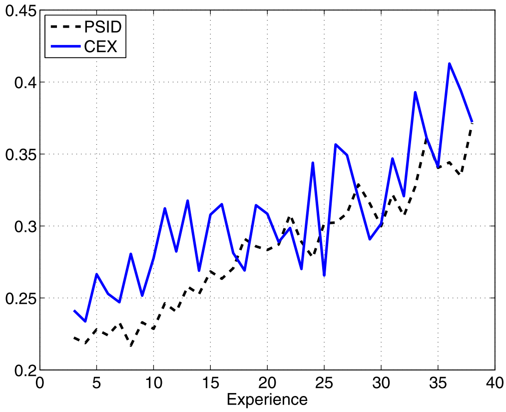
A.2 *Consumer expenditure survey*

Data source and selection. The CEX sample is drawn from the 1980–2003 waves and comes from Krueger and Perri (2006). I include only households that were interviewed for four consecutive quarters. Following Krueger and Perri (2006), I assign a household to a survey year if the fourth interview took place before April of the following year. I define the head as the CEX reference person if he is a male and as the spouse if the CEX reference person is female and the head is a spouse. This is for consistency with the PSID in which the male is always referred to as the head in male/female couples. Table 3 shows the number of observations and individuals lost at each stage of the selection process.

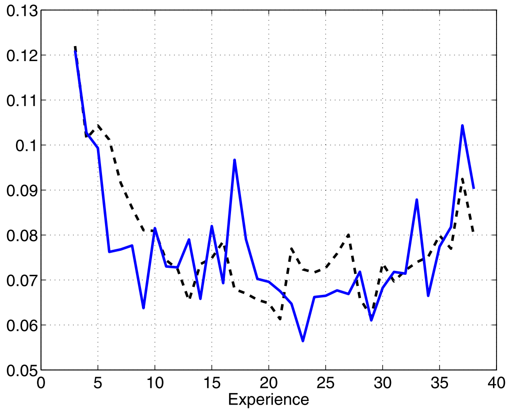
Variable definitions. The definition of nondurable consumption expenditures is discussed in Appendix C.4. This is the same definition used in Krueger and Perri (2006). I refer the reader to that paper for further details.

Table 3. Number of observations lost at each stage of the selection process.

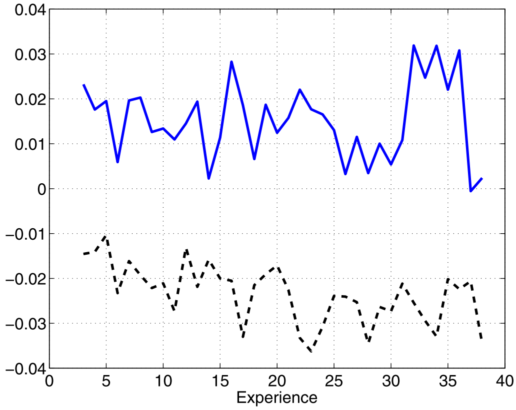
	PSID Observations	PSID Unique Households	CEX Observations
PSID SRC, 1968–2005	221,520	16,920	
CEX, 1980–2003			458,680
Heads of households	(87,262)	(4749)	
Males	(31,328)	(3876)	
Less than 4 interviews			(179,416)
Missing years ed.	(1153)	(442)	(29,037)
PSID 68, 69 waves	(4318)	(367)	
Age	(17,403)	(582)	(65,111)
Experience	(5904)	(261)	(12,614)
One primary earner	(26,744)	(1136)	(81,997)
Only 5th quarterly interview			(68,034)
Hours	(634)	(29)	(555)
Min wage	(392)	(9)	(174)
Earnings > 1M	(13)	(0)	(0)
Final sample	46,369	5469	21,742



(a) Variance log wage



(b) Variance log hours



(c) Covariance log wages, log hours

FIGURE 10. Comparison of the wage/hours distribution in CEX and PSID.

A.3 *Comparability of CEX and PSID samples*

Figure 10 shows a comparison of the life-cycle properties of the joint distribution of wages and hours in the two data sets. Both the level and the shape of the variance of wages and hours are remarkably similar in the two samples. The wage/hours correlation is negative in the PSID with a slight decrease, while it is positive in the CEX and relatively flat.

A.4 *Current population survey*

CPS data are used for the analysis of the distribution of hours over the life cycle in Section 7. My CPS sample is from the Integrated Public Use Microdata Series (IPUMS) (http://cps.ipums.org/cps/) collection of the March Outgoing Rotation Group surveys. IPUMS creates a set of individual-level and household-level sample weights. I use the individual weights in all calculations. The key variables that are used are usual hours worked per week last year (UHRSWORK) and actual weeks worked last year

(WKSWORK1). From these variables, I construct a measure of annual hours worked. I use data from the 1976–2008 surveys. For the analysis in Section 7, the only selection criterion that I impose is to restrict attention to males.

Appendix B: Construction of moments

Let z_{it}^1 and z_{it}^2 be observations on two variables for household i with potential labor market experience t. These variables may be log wages, log annual hours, or log equivalent household consumption. The covariance between any two of these variables at each t can be split into a component due to observable characteristics X and a residual component. This decomposition is given by

$$\mathrm{COV}[z^1, z^2 \mid t] = \mathrm{COV}_X(E[z^1 \mid X, t], E[z^2 \mid X, t]) + E_X(\mathrm{COV}[z^1, z^2 \mid X, t]). \tag{9}$$

In this paper, I am concerned with properties of the second, residual, component. The residual component can be further expressed as

$$E_X(\mathrm{COV}[z^1, z^2 \mid X, t]) = E\big[\varepsilon^{z^1}\varepsilon^{z^2} \mid t\big], \tag{10}$$

where $\varepsilon_{it}^z = z_{it} - E[z \mid X_{it}, t]$.

First moments. The set of variables contained in X_{it} is assumed to be the current year, education, and race. Note that this implies that the information set defined by (X_{it}, t) also implicitly includes age and cohort (year of birth). I construct an estimate $\hat{\varepsilon}_{it}^z$ of ε_{it}^z as the residual from a regression of z_{it} on a full set of experience dummies, three race dummies, and four education dummies interacted with either year or cohort dummies. Mean life-cycle profiles are those described by the coefficients on the experience dummies in this regression. For future reference in Appendix G, I note that it is possible to write these coefficients in the form of a moment $E[\tilde{z}_{it} \mid t]$, where $\tilde{z}_{it}$ is a linear function of (z_{it}, X_{it}). For data on consumption, I focus only on $E[\tilde{z}_{it} \mid t] - E[\tilde{z}_{it} \mid 3]$.

Second moments. Let x_{it} denote the product of the residuals for a relevant pair of variables, (z_{it}^1, z_{it}^2), that is, $x_{it} = \varepsilon_{it}^{z^1}\varepsilon_{it}^{z^2}$. Its estimate is given by $\hat{x}_{it} = \hat{\varepsilon}_{it}^{z^1}\hat{\varepsilon}_{it}^{z^2}$. The moment defined in (10) can be written as $E[x_{it} \mid t]$. A consistent estimate for this expectation, with large n asymptotics, is given by

$$a_n^{x,t} \equiv \frac{1}{n^{x,t}} \sum_{i=1}^{n^{x,t}} \hat{x}_{it} \to E[x_{it} \mid t], \tag{11}$$

where $n^{x,t}$ refers to the number of observations of variables (z^1, z^2) at experience level t.

The data plotted in Figures 2, 6, and 7 refer to $a_n^{x,t}$ modified to control for year effects (see below). The confidence intervals are constructed by bootstrap with 250 bootstrap repetitions. For moments derived from PSID data, the bootstrap is clustered by individual to account for the fact that the PSID follows the same people over time, so consecutive observations may not be independent. For moments derived from the CEX, the bootstrap is stratified by experience level. Since the CEX is a cross-sectional survey, clustering is not required. The confidence intervals account for additional estimation error induced by the use of residuals from the first-stage regression.

This description has not explained how identification assumptions for dealing with age, year, and cohort effects are imposed. Appendix C.3 below describes how this procedure is modified to allow for year or cohort effects in the second moments by reweighting the summation term in (11).

Appendix C: Alternative empirical choices

C.1 *Selection of single earner households*

Figure 11 compares the life-cycle profiles of the first and second moments of log wages, log hours, and log consumption across three different samples. The sample labelled Single is the sample adopted as the baseline sample in the main text: households with either a single male earner or where the second earner earns less than half the earnings of the primary earner. The sample labelled Married is the complement of the Single sample: households with two earners. The sample labelled Full is the combined sample. Along all dimensions, the life-cycle profiles have essentially the same shape. Thus none of the structural parameter estimates or economic implications is significantly affected by this choice.

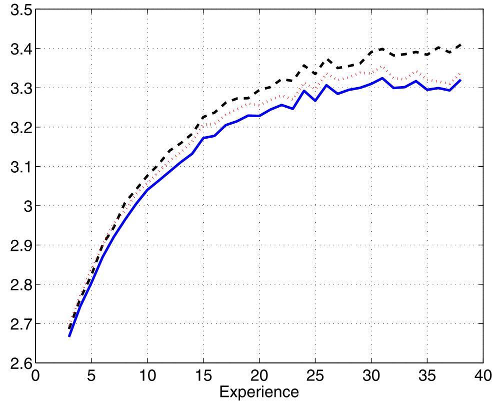
C.2 *Definition of life cycle*

For the reasons discussed in Section 3, in the main body of the paper I focus on potential labor market experience (age minus years of education minus 6) as the variable that defines the life cycle. In this appendix, I provide evidence that none of the main results of the paper would have been affected had I instead used age to define the life cycle. Figure 12 compares the life-cycle profiles of the first and second moments of log wages, log hours and log consumption across these two definitions of the life cycle. As before, potential labor market experience is plotted between 3 and 38, resulting in 36 data points. To enable comparison across the two approaches, age is plotted between 22 and 57, which also generates 36 data points. All figures are generated from the same underlying sample of observations. The two sets of plots differ only in how they assign observations to points in the life cycle. For the sake of clarity, the plots based on age are with respect to an x-axis that corresponds to age 19. Figure 12 shows clearly that all the life-cycle profiles are extremely similar, regardless of which definition is adoption.

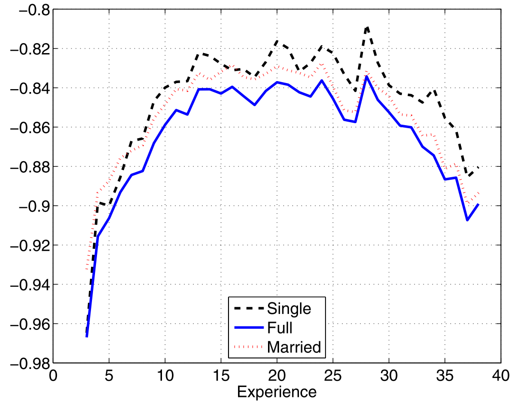
C.3 *Controlling for year/cohort effects*

Consider a squared residual variable x_{it} as defined in Appendix B. Recall that t refers to potential labor market experience (the age variable) and that x_{it} is unaffected by whether the conditional means in the first stage are based on year or cohort, since together with education and age, each partitions the observations in the same way. Let y refer to year and let k refer to cohort. By construction, $k = y - t$.

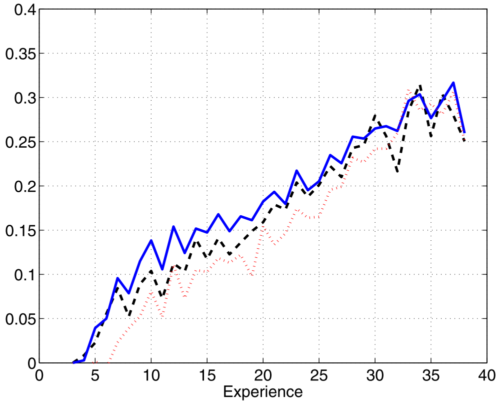
In every (t, y) cell (or equivalently (t, k)), a conditional moment of x_{it} can be computed, $E[x \mid t, y] = E[x \mid t, k]$. This is a two-dimensional function. Now, since the model



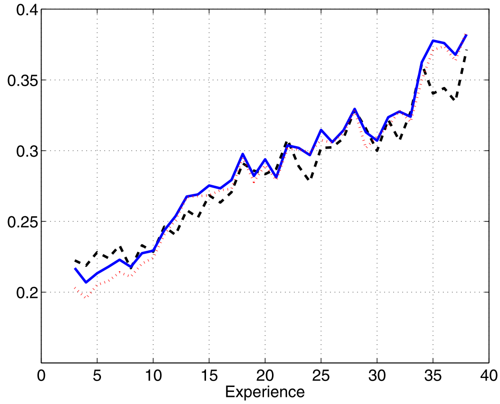
(a) Mean log wage



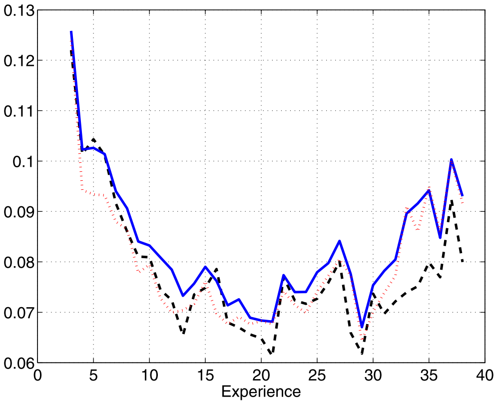
(b) Mean log hours



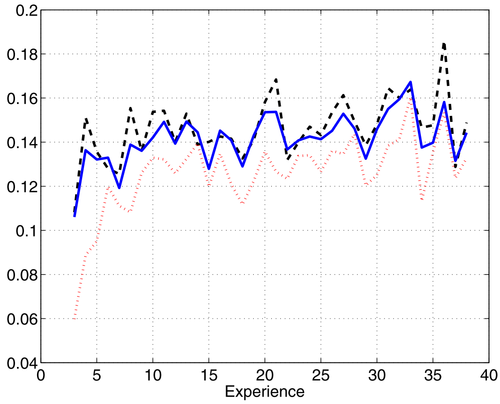
(c) Mean log consumption



(d) Variance log wage

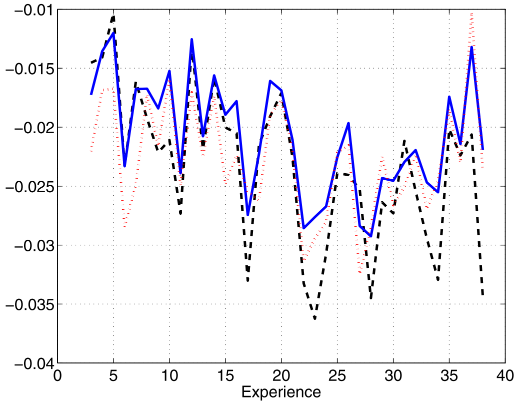


(e) Variance log hours

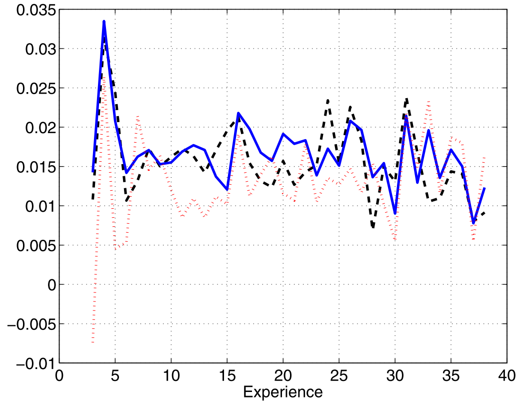


(f) Variance log consumption

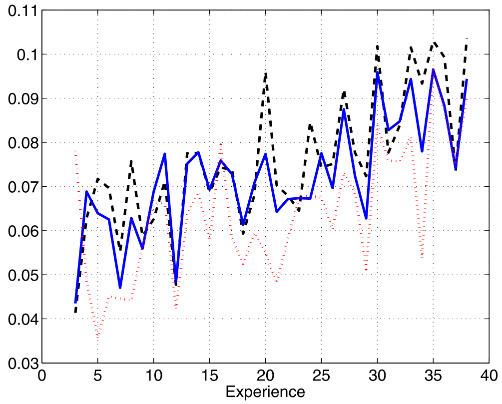
FIGURE 11. Sensitivity of life-cycle profiles to selection of households with a single primary earner.



(g) Covariance log wage, log hours



(h) Covariance log hours, log consumption



(i) Covariance log wages, log consumption

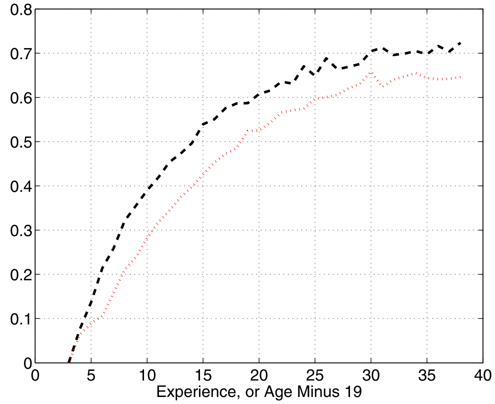
FIGURE 11. *Continued.*

is stationary, it does not distinguish between comparisons across ages that hold cohorts (k) fixed and comparisons that hold years (y) fixed. But since the world is nonstationary, these two comparisons are indeed different in the data. To confront the model with the data, it is necessary to project the two-dimensional function $E[x \mid t, y]$ onto a one-dimensional subspace to generate the function $E[x \mid t]$. The essence of the identification problem is that there are many ways to do this projection. I consider two such projections.

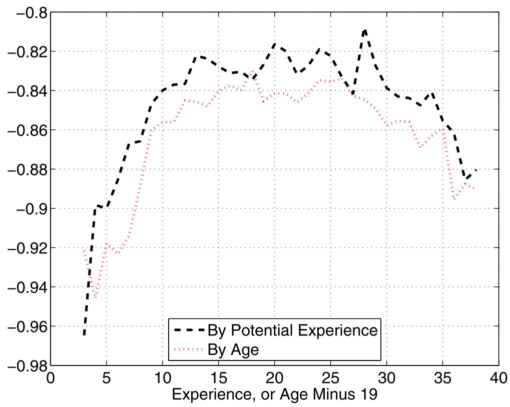
The first approach, which I call *time effects*, computes $E[x \mid t]$ as

$$E^y[x \mid t] = \frac{1}{Y} \sum_{y=1}^{Y} E[x \mid t, y],$$

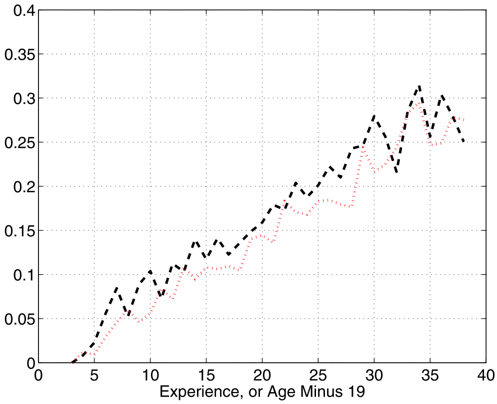
where Y is the number of years for which there are data. This approach weights the conditional moments in each year equally. This is equivalent to studying the coefficients on the age effects in a separable model with a full set of dummy variables for age and years.



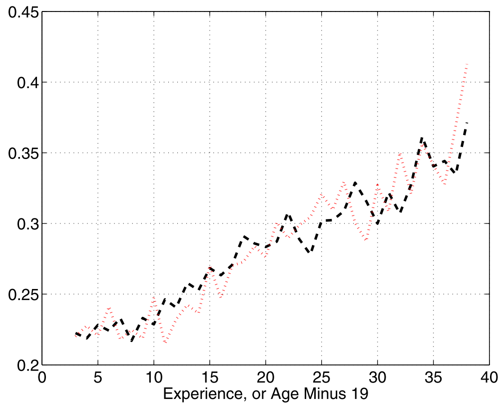
(a) Mean log wage



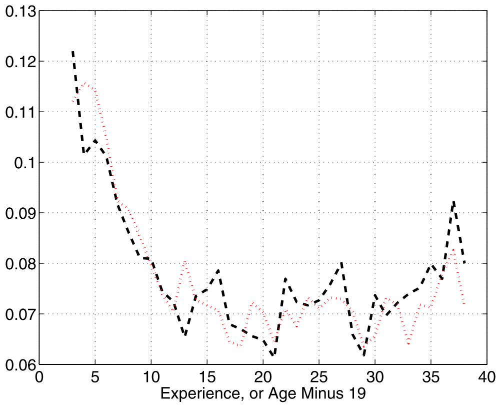
(b) Mean log hours



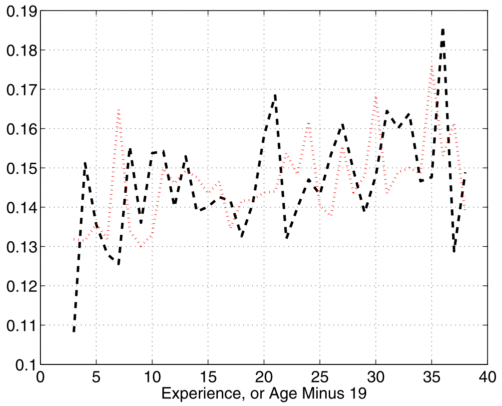
(c) Mean log consumption



(d) Variance log wage

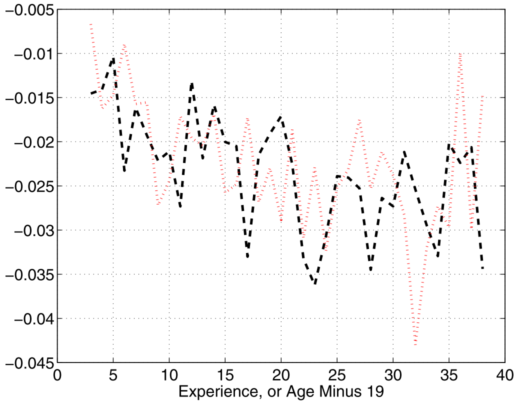


(e) Variance log hours

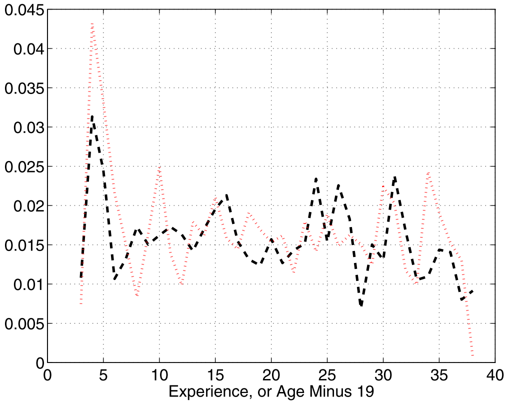


(f) Variance log consumption

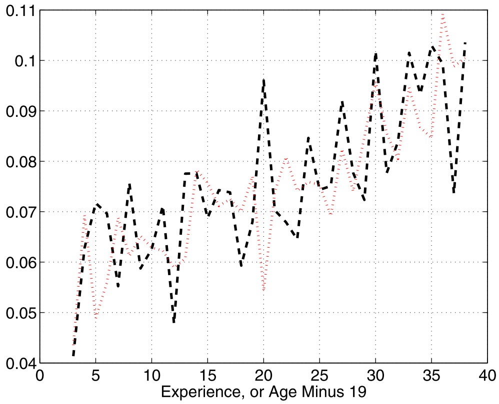
FIGURE 12. Sensitivity of life-cycle profiles to the definition of the life cycle: age versus potential experience.



(g) Covariance log wage, log hours



(h) Covariance log hours, log consumption



(i) Covariance log wages, log consumption

Figure 12. *Continued.*

The second approach, which I call *cohort effects*, computes $E[x \mid t]$ as

$$E^k[x \mid t] = \frac{1}{K(t)} \sum_k E[x \mid t, k],$$

where $K(t)$ is the number of cohorts in the sample of individuals with experience level t. This approach weights the conditional moments from each cohort equally.

To estimate these moments, I appeal to a law of large numbers and replace expectations with sample averages. This implies that each formulation can be expressed as a reweighting version of the sum in (11),

$$\hat{E}^y[x \mid t] = \frac{1}{n^{x,t}} \sum_{i=1}^{n^{x,t}} \frac{n^{x,t}}{YN(t, y)} x_{it},$$

$$\hat{E}^k[x \mid t] = \frac{1}{n^{x,t}} \sum_{i=1}^{n^{x,t}} \frac{n^{x,t}}{K(t)N(t, k)} x_{it},$$

where $N(t, y) = N(t, k)$ is the number of observations on x from individuals with experience level t in year y of cohort $k = y - t$.

I also consider a third approach, which I call *raw*, that estimates $E[x \mid t]$ as

$$\hat{E}^r[x \mid t] = \frac{1}{n^{x,t}} \sum_{i=1}^{n^{x,t}} x_{it}.$$

This is the estimate that would result from taking the sample average of x_{it} at each t. Note that because the CEX and the PSID are random samples with respect to age, they have the property that $N(t, y) \approx n^{x,t}$. Hence the time effects approach and the raw approach will give very similar results.

Figure 13 displays the life-cycle profiles of the first and second moments of log wages, log consumption, and log hours for the three different weighting functions. The dashed lines correspond to the time effects approach. These are the data used in the main body of the paper. The solid line corresponds to the cohort effects approach. The dotted line reports the raw approach.

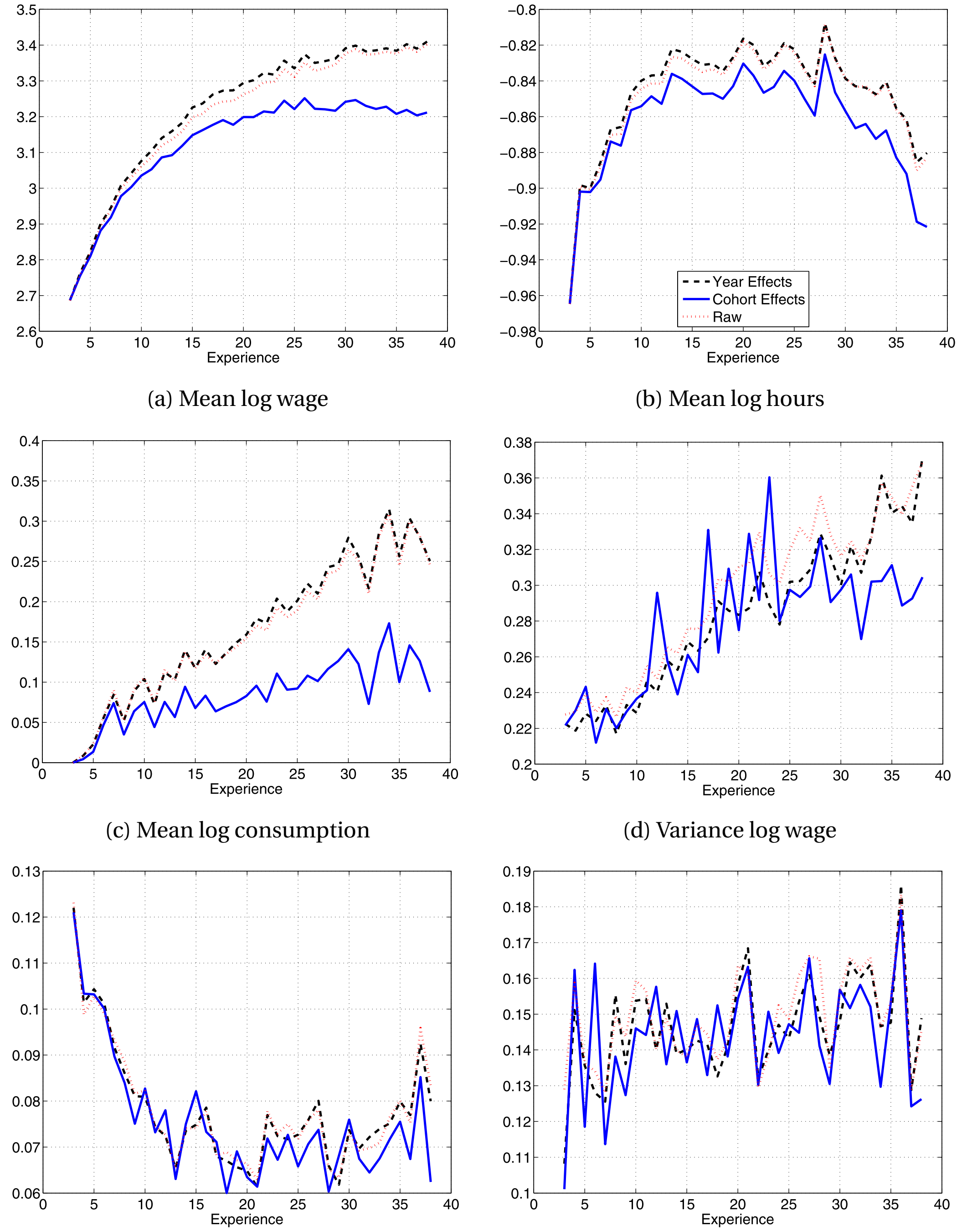
Of these life-cycle profiles, only three are sensitive to the choice of normalization: (i) mean wages, (ii) mean consumption, and (iii) variance of wages over the final 10 years. All other moments look almost identical, regardless of whether one takes a cohort view or year view of the data. Importantly, all of the moments that are needed for the arguments in Section 6 regarding failure of the benchmark model to generate declining inequality in hours at young ages are not affected by this choice.

C.4 *Consumption definition*

In this appendix, I examine the sensitivity of the life-cycle profiles of consumption to alternative definitions of consumption. In the main body of the paper, I focus on nondurable expenditures only. I adopt the definition used by Krueger and Perri (2006). This includes food, alcoholic beverages, tobacco, personal care, fuels, utilities and public services, household operations, public transportation, gasoline and motor oil, apparel, education, reading, health services, and miscellaneous expenditures. Figure 14 compares the consumption profiles using this definition to two other broader definitions.

The first alternative definition, which I label Nondurables plus services from durables adds entertainment, household equipment, other lodging expenses, other vehicle expenses, rented dwellings, imputed services from owned primary residence, and imputed services from vehicles. The second alternative definition, which I label Total expenditure replaces the imputed services with purchases of dwellings and purchases of vehicles.

Figure 14 shows that including imputed services from durables has no noticeable effect on the rise in mean consumption over the life cycle, and has almost no effect on the level or shape of the variance of consumption over the life cycle. However, including housing and vehicle purchases does have an effect: this generates a steeper rise in mean consumption and a higher cross-sectional variance of consumption at all ages.

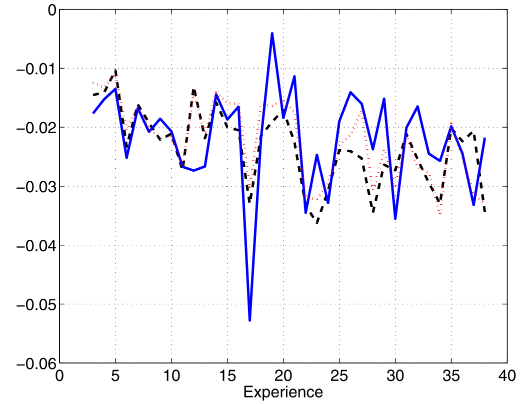


(a) Mean log wage (b) Mean log hours

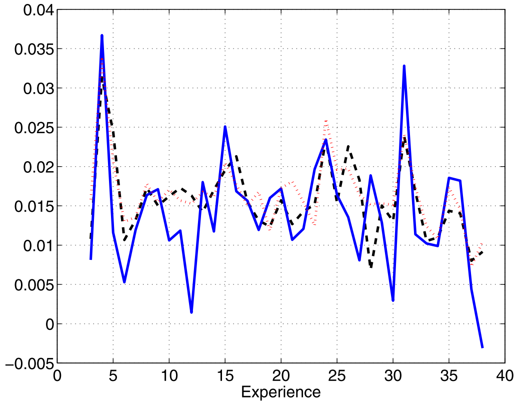
(c) Mean log consumption (d) Variance log wage

(e) Variance log hours (f) Variance log consumption

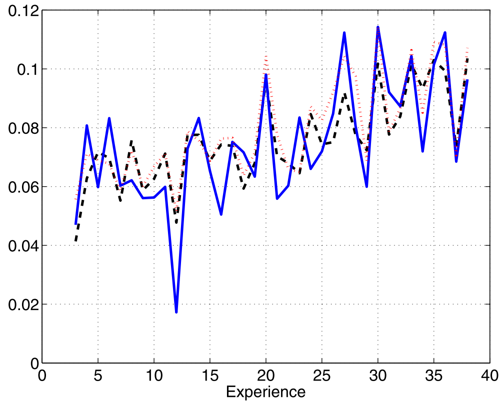
Figure 13. Sensitivity of life-cycle profiles to choice of normalization for cohort/year effects.



(g) Covariance log wage, log hours

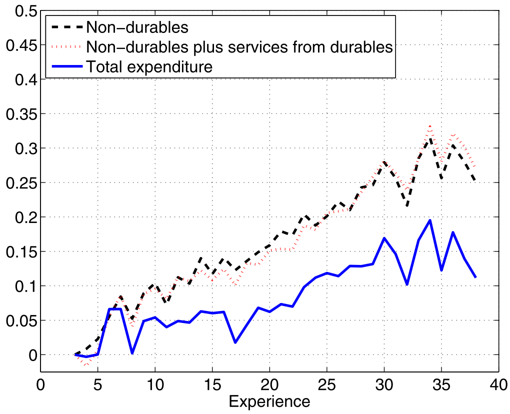


(h) Covariance log hours, log consumption

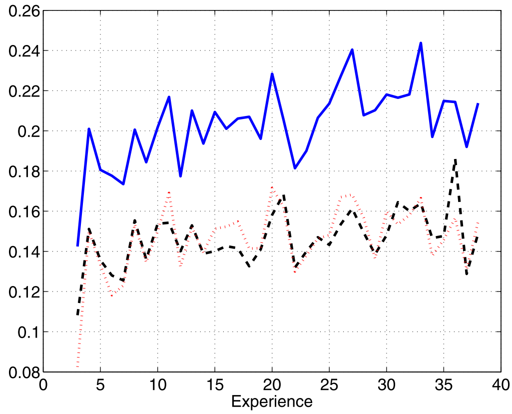


(i) Covariance log wages, log consumption

FIGURE 13. *Continued.*

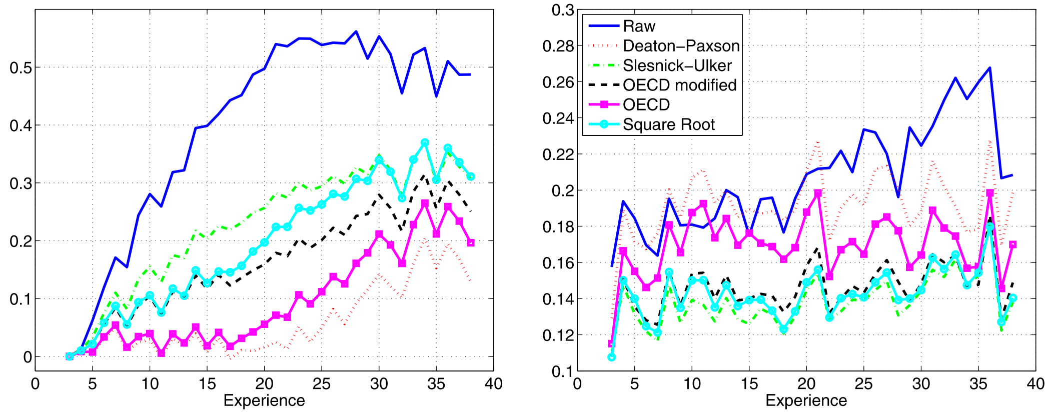


(a) Mean log consumption



(b) Variance log consumption

FIGURE 14. Sensitivity of consumption profiles to consumption definition.



(a) Mean log consumption

(b) Variance log consumption

FIGURE 15. Sensitivity of consumption profiles to equivalence scale.

C.5 *Choice of equivalence scale*

Figure 15 shows that the life-cycle profile for both the mean and the variance of log consumption is sensitive to the choice of household equivalence scales. The definitions of the various equivalence scales are as follows.

Deaton–Paxson: This is the equivalence scale used in Deaton and Paxson (1994). It assigns 1 for each adult in the household and 0.5 for each child.

Slesnick–Ulker: This is the equivalence scale used in Slesnick and Ulker (2004). It is equal to the square root of the Deaton–Paxson scale.

OECD: This scale, also known as the Oxford scale, assigns 1 to the first household member, 0.7 to each additional adult, and 0.5 to each additional child.

OECD-modified: This scale, first proposed by Hagenaars, De Vos, and Zaidi (1994), assigns 1 to the first household member, 0.5 to each additional adult, and 0.3 to each additional child.

Square root: This scale is equal to the square root of the number of household members, regardless of age.

There does not seem to be a consensus about which is the best equivalence scale to use. Each of these five scales has been used in parts of the literature.

APPENDIX D: COMPOSITION EFFECTS IN HOURS VARIANCE

In this appendix, I describe the procedure used to construct Figure 3(b) and 3(c). For any random variable Y, we can decompose its variance, $V(Y)$, into within-group and between-group components, where groups are define by the variable X:

$$V(Y) = V_X[E(Y \mid X)] + E_X[V(Y \mid X)]. \tag{12}$$

The second term in (12) is the within-group variance. Denote its value at age t by $V_t^W(Y)$. The change in $V_t^W(Y)$ can be further decomposed as

$$\Delta V_t^W(Y) = \int_X \Delta V_t^W(Y \mid X)\, dF_t(X) + \int_X V_{t-1}(Y \mid X) \Delta dF_t(X). \tag{13}$$

The first term in (13) is the true change in the within-group variance: the change that would occur if the distribution of conditioning variables (X) stayed constant with age. The second term in (13) is the composition effect: the change that is due to differences in the distribution of X with age.

To construct the path for $V_t^w(Y)$ net of the composition effect, I reconstruct the sequence of $\Delta V_t^W(Y)$ using only the first term in (13). The set of X variables that I consider is discrete: it includes employment status, labor force status, education status, marital status, number of children, home ownership, and race. In each cell defined by these variables, I first construct an estimate of $V_{t-1}(Y \mid X)$. I then estimate the second term in (13) as the difference between the expectation of this variance over X using the age t and age $t-1$ distributions of X. The first term in (13) is then obtained as a residual.

Appendix E: Wage process

E.1 *Identification of parameters*

Let ω_{it} be the estimated log wage residual from the first-stage regression, that is, $\omega_{it} \equiv \varepsilon_{it}^w$. The benchmark specification for the statistical process governing ω_{it} is

$$\omega_{it} = \kappa_t + \alpha_i + z_{it} + \varepsilon_{it},$$

$$z_{it} = \rho z_{it-1} + \eta_{it},$$

$$z_{i0} = 0,$$

where $E[\varepsilon_{it}] = E[\eta_{it}] = E[\alpha_i] = 0$, and $V[\eta_{it}] = v_\eta$, $V[\varepsilon_{it}] = v_{\varepsilon t}$, and $V[\alpha_i] = v_\alpha$. Denote an element of the autocovariance function of ω_{it} as $\sigma_{t,t+j} \equiv \mathrm{COV}(\omega_{i,t}, \omega_{i,t+j})$. The autocovariance function for this process is then given by

$$\sigma_{tt} = v_\alpha + \frac{1-\rho^{2t}}{1-\rho^2} v_\eta + v_{\varepsilon t},$$

$$\sigma_{ts} = v_\alpha + \rho^{(s-t)} \frac{1-\rho^{2t}}{1-\rho^2} v_\eta \quad \text{for } s > t.$$

Assume that ω_{it} is observed for $t = 1, T$ with $T \geq 4$. The following subset of the available moment conditions uniquely identifies $\{\rho, v_\alpha, v_\eta, \}$:

$$\frac{\sigma_{14} - \sigma_{13}}{\sigma_{13} - \sigma_{12}} = \rho,$$

$$\sigma_{13} - \sigma_{12} = \rho(\rho - 1) v_\eta,$$

$$\sigma_{12} = v_\alpha + \rho v_\eta.$$

Given identification of these parameters, $\{v_{\varepsilon t}\}_{t=1}^{T}$ is identified from $\{\sigma_{tt}\}_{t=1}^{T}$. The deterministic earnings profile is identified directly from $E[\omega_{it} \mid t]$.

E.2 *Estimation of wage process*

To estimate the wage process, I first construct the empirical analogue of the autocovariance function by year. That is, I construct an estimate of the autocovariances $\mathrm{COV}(\omega_{i,t}, \omega_{i,t-j})$ in each experience/year cell. Note that since experience/year and experience/cohort partition the experience/cohort/year space in exactly the same way, it is irrelevant whether this is done in terms of year or cohort. Note also that, given that the wage process does not have year or cohort effects, it is not strictly necessary to construct autocovariance functions for each year. However, it has no impact on the estimation, yet it permits me to compare my estimates with what would have obtained if I had allowed for year or cohort effects in the variances of the wage shocks. I use a maximum of 25 lags in each cell and only retain moments that were constructed with at least 30 observations. The results are not sensitive to alternative choices for either of these values.

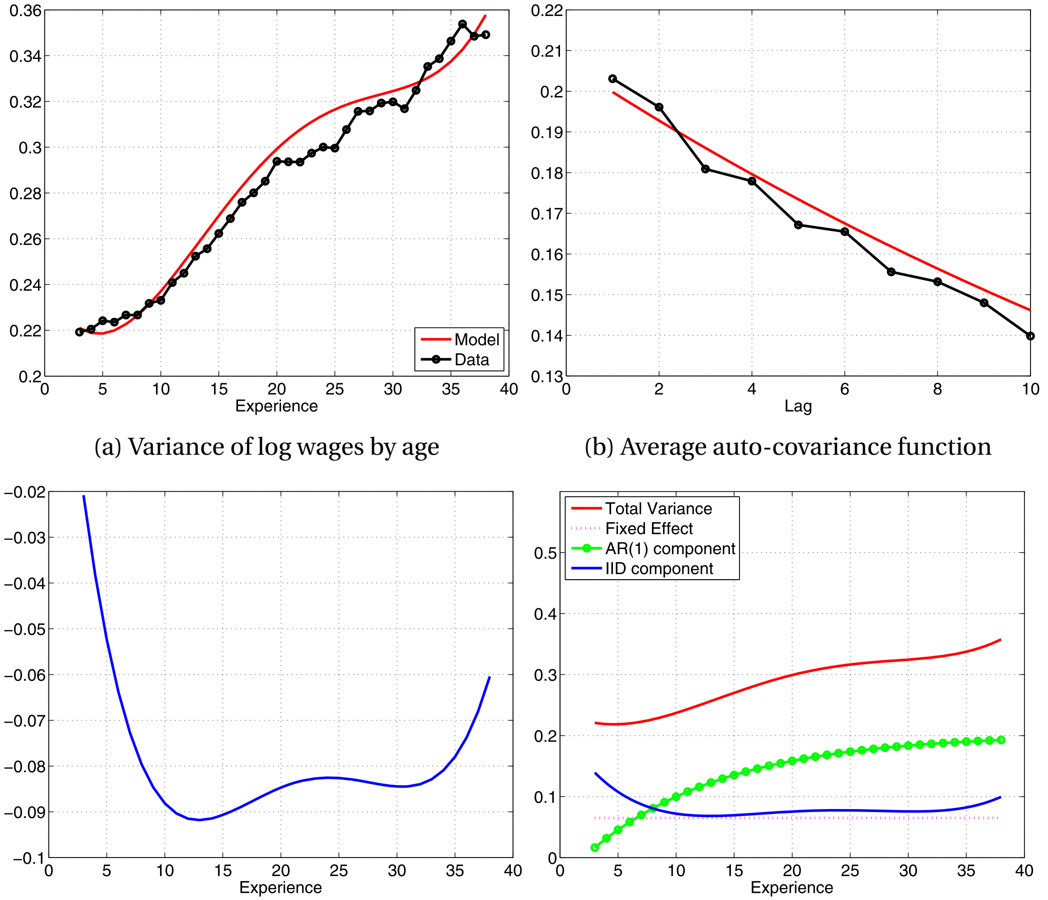
I parametrize the age effects in the transitory variance using a quartic polynomial. This turns out to be flexible enough to match the age patterns in the autocovariance function well.

The parameters are estimated using the generalized method of moments with a weighting matrix given by $n^{-1/2}$, where n is the number of observations used in the construction of each corresponding sample moment. Standard errors are obtained by bootstrap with 250 repetitions.

The parameter estimates from the benchmark specification are reported in Table 4. The fit of the wage model is displayed in Figure 16 for the model with year effects. The model with cohort effects generates almost identical figures. Figure 16(a) and (b) shows that the model provides a good fit to the variance of wages over the life cycle and the average autocovariance function. The estimated experience effects in the transitory variance of wages are displayed in Figure 16(c) and a decomposition of the variance of wages into its various components can be found in Figure 16(d).

TABLE 4. Parameter estimates.

			With Time Effects in $(\nu_\rho, \nu_\varepsilon)$	
	Year Effects	Cohort Effects	Year Effects	Cohort Effects
Estimated parameters				
ρ	0.958	0.948	0.970	0.964
	(0.945, 0.969)	(0.921, 0.958)	(0.958, 0.988)	(0.940, 0.980)
v_α	0.065	0.059	0.066	0.059
	(0.048, 0.088)	(0.036, 0.069)	(0.047, 0.087)	(0.041, 0.070)
v_ρ	0.017	0.018	0.016	0.017
	(0.013, 0.021)	(0.015, 0.023)	(0.010, 0.020)	(0.012, 0.022)
$\overline{v_\varepsilon}$	0.081	0.081	0.080	0.081
	(0.074, 0.087)	(0.072, 0.090)	(0.067, 0.090)	(0.074, 0.093)



(a) Variance of log wages by age

(b) Average auto-covariance function

(c) Estimated experience effects in transitory wage shocks

(d) Decomposition of wage variance

FIGURE 16. Model fit for estimated wage process (year effects).

E.3 *Alternate specifications*

I have also considered numerous modifications to the baseline specification. These include not allowing for age effects in the transitory variance; allowing for time effects in the transitory and persistent variances as well as the "price" of the fixed effect; allowing for profile heterogeneity in the manner advocated by Guvenen (2009); using alternative weighting matrices and different sets of lags for the targeted autocovariance matrix; using alternative sample choices and year/experience restrictions. The possible formulations are far too numerous to report. Table 4 reports estimates that allow for year effects in the variances of the persistent and transitory shocks, rather than age effects in the transitory shocks. None of the parameter estimates is significantly affected. All other results are available from the author on request.

Appendix F: Further details of the model

F.1 *Social Security system and calibration*

Pension benefits are calibrated to approximate the redistribution implicit in the U.S. Social Security system. In the United States, Social Security benefits are a function of average qualifying earnings over the highest 30 years. In the model, I approximate average earnings by using a function of the fixed component of wages. In a previous version of the paper, I allowed pension benefits to also vary with the final persistent component of wages. However, as pointed out by an anonymous referee, that formulation made pension wealth overly sensitive to wage realizations in the last few working years, which is the opposite of what happens in reality. The approximation that I impose assumes that agents work average hours in every year and earn average wages conditional on their fixed effect in every year.

To match the progressivity of the U.S. Social Security system, I specify that benefits are equal to 90% of approximated average earnings up to a given bend point, 32% from this first bend point to a second bend point, and 15% beyond that. The two bend points are set at, respectively, 0.18 and 1.10 times cross-sectional average gross earnings, based on U.S. legislation and individual earnings data for 1990. Benefits are then scaled proportionately so that a worker earning average labor income each year is entitled to a replacement rate of 34%. In October 2009, average Social Security benefits for those aged 65 years and older was \$12,938, while average labor income was \$41,335. This implies an average replacement rate of 34%. See http://www.ssa.gov/policy/docs/quickfacts/stat_snapshot/.

F.2 *Progressive labor income taxes and calibration*

I assume the following functional form for labor income taxes:

$$\varsigma(y) = \tau_0 + y - \tau_1 \frac{y^{\tau_2+1}}{\tau_2+1}. \tag{14}$$

This specification implies that 1 minus the marginal tax rate is given by

$$1 - \varsigma'(y) = \tau_1 y^{\tau_2}, \tag{15}$$

$$\log(1 - \varsigma'(y)) = \log \tau_1 + \tau_2 \log y. \tag{16}$$

Hence $\log(1 - \varsigma'(y))$ is linear in log labor earnings. This specification is similar to the one used by Guvenen, Kuruscu, and Ozkan (2009) and is the same as that in Heathcote, Storesletten, and Violante (2010b). Both of these papers provide evidence that 1 minus the marginal tax rate is approximately log linear in earnings for the United States.

To estimate (τ_1, τ_2), I regress marginal tax rates for each individual in the baseline sample on labor earnings. Marginal tax rates are calculated using the National Bureau of Economic Research's TAXSIM program. The estimated parameter values are $\hat{\tau}_1 = 0.63768$ and $\hat{\tau}_2 = -0.13615$ with an R^2 of 0.42. I set ς_0 to the value that equates the actual average tax rate in the sample (as computed by TAXSIM) to that implied by (14). A regression of the actual tax liability on the predicted tax liability yields an R^2 of 0.96.

F.3 *Numerical solution*

The model is solved by backward induction using the method of endogenous grid points (Carroll (2006)), extended to allow for a labor supply decision. The grid for financial assets contains 30 points, spaced such that there are more points closer to the borrowing limit. Policy rules are approximated with piecewise linear functions on this grid. All other state variables are discrete. The three components of the wage process are approximated using a discrete-state Markov chain, with an age-varying state. Values and transition probabilities are chosen to match the age-varying unconditional variance and dependence structure of each component to that implied by the continuous process. I use 5 points for the fixed effects, 5 points for the transitory shocks, and 11 points for the persistent component. The grid for preference heterogeneity in the disutility of working contains 3 points. None of the results is sensitive to increasing the size of the grids in any dimension. Cross-sectional distributions are obtained by simulation with 50,000 simulations at each age.

For minimization of the simulated method of moments objective function, I use the derivative-free least squares algorithm of Zhang, Conn, and Scheinberg (2010). This is a trust region based approach that works by sequentially building up a series of smooth approximations to each of the moment conditions. It works well with nonsmooth objective functions and requires far fewer function evaluations than other derivative-free algorithms, such as simplex methods or simulated annealing. However, the algorithm is only guaranteed to converge to a local minimum. Hence I conduct an extensive search of the parameter space using multiple restarts of the algorithm.

Appendix G: Estimation

G.1 *Formal description of estimator*

Let $\theta \in \Theta$ denote the $K \times 1$ vector of structural parameters to be estimated. I assume that Θ can be restricted to a convex and compact subset of $\mathbb{R}^K$, and that the true parameter θ_0 lies on the interior of Θ. Let x_{it} denote the variables that are used to form either a first moment or second moment as defined in Appendix B. For first moments, x_{it} reflects a coefficient on an experience dummy from the first-stage regression (or deviation from the coefficient at experience level 3 in the case of consumption data). For second moments, x_{it} reflects the product of a pair of residuals from the first-stage regression. There are $M > K$ such transformed variables,

For each x and t, define the corresponding observation for a simulation $r \in R$ from the structural model with parameter vector θ as $\alpha_r^{x,t}(\theta)$. A standard law of large numbers ensures that $\frac{1}{R}\sum_{r=1}^{R}\alpha_r^{x,t}(\theta) \to E[\alpha_r^{x,t}(\theta)] \equiv \alpha^{x,t}(\theta)$, the corresponding moment implied by the structural model. The identifying assumption that underlies estimation is

$$E[x_{it} - \alpha^{x,t}(\theta)] = 0 \quad \text{if and only if} \quad \theta = \theta_0.$$

The simulated method of moments estimator is defined as

$$\begin{aligned} \hat{\theta}_n &= \arg\max Q_n(\theta), \\ Q_n(\theta) &= -\frac{1}{2}\hat{g}_n(\theta)\hat{\Omega}_n^{-1}\hat{g}_n(\theta), \end{aligned} \tag{17}$$

where $\hat{\Omega}_n$ is a positive semidefinite matrix and $\hat{g}_n(\theta)$ is an $M \times 1$ vector with elements given by

$$\hat{g}^{x,t}(\theta) \equiv \frac{1}{n^{x,t}} \sum_{i=1}^{n^{x,t}} \left(x_{it} - \frac{1}{R} \sum_{r=1}^{R} \alpha_r^{x,t}(\theta) \right).$$

Provided that $R \to \infty$ at a rate that is faster than $(n^{x,t})^{0.25}$ for all (x, t), simulation error in the calculation of $\alpha^{x,t}(\theta)$ does not impact consistency of $\hat{\theta}_n$, confidence intervals for $\hat{\theta}_n$, or construction of the test of overidentifying restrictions. With standard assumptions, it is straightforward to show that $\theta_n \to \theta$.

Define the covariance matrix of the moment conditions Ω as

$$\Omega \equiv E\big[(x_{it} - \alpha^{x,t}(\theta_0))(x_{it} - \alpha^{x,t}(\theta_0))'\big].$$

If $\hat{\Omega}_n$ is chosen such that $\hat{\Omega}_n \to \Omega$, then one can show that under the null hypothesis that $E[x_{it} - \alpha^{x,t}(\theta_0)] = 0$, the statistic $\hat{J}_n$ converges to a chi-squared distribution with $M - K$ degrees of freedom, where

$$\hat{J}_n = \left[\frac{1}{\sqrt{n^{x,t}}} \sum_{i=1}^{n^{x,t}} \left(x_{it} - \frac{1}{R} \sum_{r=1}^{R} \alpha_r^{x,t}(\theta) \right) \right] \hat{\Omega}_n^{-1} \left[\frac{1}{\sqrt{n^{x,t}}} \sum_{i=1}^{n^{x,t}} \left(x_{it} - \frac{1}{R} \sum_{r=1}^{R} \alpha_r^{x,t}(\theta) \right) \right]'.$$

A positive semidefinite matrix $\hat{\Omega}_n$ that has this property is the sample variance–covariance matrix of x_{it}. In estimation I use a diagonal approximation to $\hat{\Omega}_n$ that sets the off-diagonal elements to zero.

G.2 *Confidence intervals*

The 95% confidence intervals are constructed by bootstrap. Based on arguments in Hall (1986) I use 39 repetitions. The following steps are taken:

Step 1. Thirty-nine bootstrap samples are drawn from the PSID and the CEX as described in Appendix B.

Step 2. In each bootstrap sample, I run the first stage regressions and construct the relevant moments as described in Appendix B.

Step 3. In each bootstrap sample, I reestimate the parameters that govern the wage process.

Step 4. In each bootstrap sample, I reestimate the structural parameters by maximizing (17).

Step 5. For each structural parameter, the 95% confidence interval is defined by the lowest and highest parameter estimates across the bootstrap repetitions.

Appendix H: First-order condition with general preference specifications

H.1 *No heterogeneity*

I start by considering the case with no cross-sectional heterogeneity in preferences. Consider an agent with period utility given by $u(c, H)$. The intratemporal first-order condition is given by

$$w_{it} u_c(c_{it}, H_{it}) = u_H(c_{it}, H_{it}).$$

Taking logs followed by a first-order Taylor expansion around mean log consumption and mean log hours at each age yields the approximation

$$\log w_{it} + (f_c - g_c)\log c_{it} \approx (f_H - g_H)\log H_{it} + \text{constant}, \tag{18}$$

where

$$f_c \equiv \left.\frac{\partial \log u_c(c, H)}{\partial \log c}\right|_{(\log c, \log H) = E[(\log c, \log H)|t]}$$

is the elasticity of the marginal utility of consumption with respect to consumption, evaluated at the age-specific mean levels of log consumption and log hours; f_H is the corresponding elasticity with respect to hours worked; and g_c and g_H are the analogous elasticities of the marginal disutility of hours worked.

Taking cross-sectional variances and first difference of (18) yields a generalized version of equation (3):

$$\begin{aligned}(f_H - g_H)^2 \Delta V(\log H_{it}) = \Delta V(\log w_{it}) + (f_c - g_c)^2 \Delta V(\log c_{it})\\ + 2(f_c - g_c)\Delta \operatorname{COV}(\log c_{it}, \log w_{it}).\end{aligned} \tag{19}$$

Inspection of (19) highlights why allowing for more general patterns of complementarity or substitutability between consumption and hours is unlikely to deliver a declining profile for the variance of hours while respecting the data on consumption and wages. Given the increasing profile for the variance of wages and consumption, and a mildly increasing covariance between wages and consumption, the only way to generate a decrease in the variance of hours is by setting $g_c \gg f_c$. However, as with the case discussed in Section 6, when $g_c - f_c > 2$, there is an offsetting effect through the variance of consumption. Hence the scope for using this channel is limited.

H.2 *Allowing for heterogeneity*

If we allow for cross-sectional heterogeneity in the various elasticities (f_c, f_H, g_c, g_H) and if we allow the degree of heterogeneity to vary with age, then it is possible to generate any age profile for the variance of hours. This was true even in the separable specification discussed in Section 6. A problem with pursuing this approach is that it is difficult to provide a microfoundation for such age-varying heterogeneity.

However, with heterogeneity only in the relative distaste for work compared with consumption, the argument in Section 6 carries over to much more general preference specifications. Here I illustrate this fact with constant elasticity of substitution (CES) preference. For algebraic simplicity, I allow preferences to depend on consumption (c_{it}) and the inverse of hours worked ($\frac{1}{H_{it}}$), although the same arguments hold (up to appropriate approximations) if leisure ($1 - H_{it}$) were used instead.

Assume that the period utility function is given by

$$\frac{1}{1-\gamma}\left[\frac{c_{it}^{\nu}}{1+\varphi_i}+\frac{\varphi_i\left(\frac{1}{H_{it}}\right)^{\nu}}{1+\varphi_i}\right]^{(1-\gamma)/\nu}.$$

Then the analogous expression to (3) is

$$\begin{aligned}(\nu+1)^2\Delta V(\log H_{it}) = {} & \Delta V(\log w_{it}) + (\nu-1)^2\Delta V(\log c_{it}) \\ & - 2(\nu-1)\Delta\,\mathrm{COV}(\log c_{it}, \log w_{it}) \\ & + 2(\nu-1)\,\mathrm{COV}(\Delta\log c_{it}, \log\varphi_i).\end{aligned} \tag{20}$$

TABLE 5. Parameter estimates with cohort effects. The structural parameter estimates and bootstrap confidence intervals are shown when cohort effects, rather than year effects, are allowed for in first and second moments. The estimated borrowing limit of −10.0 in column 2 reflects an estimate on the boundary of the parameter space.

	(1)	(2)	(3)
Estimated parameters			
γ	2.01	2.73	2.04
	(1.33, 5.15)	(1.54, 4.01)	(1.39, 3.30)
σ	2.67	3.91	2.78
	(1.81, 3.66)	(2.25, 10.23)	(1.45, 4.03)
CV[φ]	1.56	1.16	1.36
	(0.37, 142.15)	(0.54, 3.40)	(0.55, 9.77)
$\underline{a}$	−10.0		−0.76
	(−10.0, −0.19)		(−10.0, −0.1.0)
$\sigma^2_{\epsilon,y}$			0.003
			(0.002, 0.524)
$\sigma^2_{\epsilon,h}$			0.00
			(0.00, 0.01)
Test of overidentifying restrictions			
p-value	0.45	0.00	0.00
Calibrated parameters			
β	0.988	0.983	0.988
$E\varphi$	139.6	1809.5	145.6
Targeted moments			
$E[\log c]$	X	X	X
$E[\log h]$	X	X	X
$V[\log c]$			X
$V[\log h]$			X

It is straightforward to see from (20) that a similar argument applies.

Appendix I: Parameter estimates with cohort effects

In this appendix, I report parameter estimates from the structural model when allowing for cohort effects rather than year effects in the first and second moments. To conserve on space, I do not reproduce the corresponding figures that show the fit of the model. These figures are all available from the author on request. The parameter estimates that correspond to those in Table 1 are shown in Table 5, while the estimates that correspond to those in Table 2 are shown in Table 6.

Table 6. Parameter estimates with cohort effects. The structural parameter estimates and bootstrap confidence intervals are shown when cohort effects, rather than year effects, are allowed for in first and second moments. The estimated borrowing limit of −10.0 in column 2 reflects an estimate on the boundary of the parameter space.

	(1)	(2)	(3)	(4)
Estimated parameters				
γ	3.95	3.83	3.01	2.98
	(1.84, 7.89)	(2.24, 8.51)	(2.37, 4.72)	(2.55, 4.43)
σ	3.57	3.48	4.04	3.88
	(2.27, 7.20)	(2.33, 7.84)	(2.65, 8.12)	(3.07, 7.95)
CV[φ]	0.29	0.34	0.77	0.76
	(0.15, 0.51)	(0.24, 0.82)	(0.69, 0.94)	(0.68, 0.93)
$\underline{a}$	−0.42	−0.36	−0.37	−0.36
	(−7.08, −0.17)	(−9.35, −0.12)	(−4.17, −0.08)	(−7.01, −0.07)
$\sigma^2_{\epsilon,y}$	0.001	0.002	0.06	0.05
	(0.001, 1.48)	(0.000, 2.76)	(0.00, 49.23)	(0.01, 19.47)
$\sigma^2_{\epsilon,h}$	0.00	0.00	0.00	0.00
	(0.00, 0.01)	(0.00, 0.00)	(0.00, 0.00)	(0.0, 0.00)
χ	−0.48		−0.07	
	(−0.63, −0.14)		(−0.35, 0.08)	
$\rho_{\varphi,\overline{h}}$		0.75		0.04
		(0.39, 1.16)		(−0.07, 0.72)
Test of overidentifying restrictions				
p-value	0.02	0.08	0.00	0.00
Calibrated parameters				
β	0.984	0.983	0.985	0.985
$E\varphi$	2520.3	2151.5	1916.1	1554.4
Targeted moments				
$E[\log c]$	X	X	X	X
$E[\log h]$	X	X	X	X
$V[\log c]$	X	X	X	X
$V[\log h]$	X	X	X	X
COV[$\log w, \log h$]			X	X
COV[$\log c, \log h$]			X	X
COV[$\log w, \log c$]			X	X

References

Aguiar, M. and E. Hurst (2007), "Life-cycle prices and production." *American Economic Review*, 97 (5), 1533–1559. [471]

Aguiar, M. and E. Hurst (2008), "Deconstructing lifecycle expenditure." Working Paper, NBER. [471]

Altonji, J. and L. Segal (1996), "Small-sample bias in GMM estimation of covariance structures." *Journal of Business & Economic Statistics*, 14 (3), 353–366. [480]

Attanasio, O., J. Banks, C. Meghir, and G. Weber (1999), "Humps and bumps in lifetime consumption." *Journal of Business & Economic Statistics*, 17 (1), 22–35. [471, 475]

Attanasio, O., H. Low, and V. Sánchez-Marcos (2005), "Female labor supply as insurance against idiosyncratic risk." *Journal of the European Economic Association*, 3 (2–3), 755–764. [475]

Badel, A. and M. Huggett (2010), "Interpreting life-cycle inequality patterns as an efficient outcome: Mission impossible?" Working Paper, Georgetown University. [485, 488]

Blundell, R., M. Browning, and C. Meghir (1994), "Consumer demand and the life-cycle allocation of household expenditures." *Review of Economic Studies*, 61 (1), 57–80. [471, 475]

Blundell, R., L. Pistaferri, and B. Preston (2008), "Consumption inequality and partial insurance." *American Economic Review*, 98 (5), 1887–1921. [471]

Carroll, C. (2006), "The method of endogenous gridpoints for solving dynamic stochastic optimization problems." *Economics Letters*, 91 (3), 312–320. [518]

Deaton, A. and C. Paxson (1994), "Intertemporal choice and inequality." *Journal of Political Economy*, 102 (3), 437–467. [471, 478, 513]

Domeij, D. and J. Heathcote (2004), "On the distributional effects of reducing capital taxes." *International Economic Review*, 45 (2), 523–554. [481]

Erosa, A., L. Fuster, and G. Kambourov (2011), "Towards a micro-founded theory of aggregate labor supply." Unpublished Manuscript, University of Toronto. [473, 488]

French, E. (2005), "The effects of health, wealth, and wages on labour supply and retirement behaviour." *Review of Economic Studies*, 72 (2), 395–427. [474]

Gourinchas, P. and J. Parker (2002), "Consumption over the life cycle." *Econometrica*, 70 (1), 47–89. [471, 474]

Guvenen, F. (2007), "Learning your earning: are labor income shocks really very persistent?" *American Economic Review*, 97 (3), 687–712. [478]

Guvenen, F. (2009), "An empirical investigation of labor income processes." *Review of Economic Dynamics*, 12 (1), 58–79. [516]

Guvenen, F., B. Kuruscu, and S. Ozkan (2009), "Taxation of human capital and wage inequality: A cross-country analysis." Working Paper, Federal Reserve Bank of Minneapolis. [517]

Guvenen, F. and A. Smith (2009), "Inferring labor income risk from economic choices: An indirect inference approach." Mimeo, University of Minnesota. [474]

Hagenaars, A., K. De Vos, and M. Zaidi (1994), *Poverty Statistics in the Late 1980s: Research Based on Micro-Data*. Office for Official Publications of the European Communities, Luxembourg. [513]

Hall, P. (1986), "On the number of bootstrap simulations required to construct a confidence interval." *The Annals of Statistics*, 14, 1453–1462. [519]

Hall, R. (1971), "The measurement of quality change from vintage price data." In *Price Indexes and Quality Change*, Vol. 2, Harvard University Press, Cambridge, MA. [478]

Heathcote, J., K. Storesletten, and G. Violante (2005), "Two views of inequality over the life cycle." *Journal of the European Economic Association*, 3 (2–3), 765–775. [478]

Heathcote, J., K. Storesletten, and G. Violante (2009), "Consumption and labor supply with partial insurance: An analytical framework." Working Paper, NBER. [474]

Heathcote, J., K. Storesletten, and G. Violante (2010a), "The macroeconomic implications of rising wage inequality in the united states." *Journal of Political Economy*, 118 (4), 681–722. [474]

Heathcote, J., K. Storesletten, and G. Violante (2010b), "Redistributive taxation in a partial-insurance economy." Working Paper, Federal Reserve Bank of Minneapolis. [474, 517]

Heckman, J. (1974), "Life cycle consumption and labor supply: An explanation of the relationship between income and consumption over the life cycle." *American Economic Review*, 64 (1), 188–194. [473]

Heckman, J. and T. MaCurdy (1980), "A life cycle model of female labour supply." *The Review of Economic Studies*, 47 (1), 47–74. [473, 475]

Heckman, J. and R. Robb (1985), "Using longitudinal data to estimate age, period and cohort effects in earnings equations." In *Cohort Analysis in Social Research: Beyond the Identification Problem*, 137–150, Springer, New York. [478]

Imai, S. and M. Keane (2004), "Intertemporal labor supply and human capital accumulation." *International Economic Review*, 45 (2), 601–641. [474]

Kaplan, G. and G. Violante (2010), "How much consumption insurance beyond self-insurance?" *American Economic Journal: Macroeconomics*, 2 (4), 1–37. [471]

Kocherlakota, N. (2010), *The New Dynamic Public Finance*. Princeton University Press, Princeton, NJ. [474]

Krueger, D. and F. Perri (2006), "Does income inequality lead to consumption inequality?" *Review of Economic Studies*, 73 (1), 163–193. [502, 510]

Lise, J. (2010), "On-the-job search and precautionary savings: Theory and empirics of earnings and wealth inequality." Working paper, University College London. [473]

Low, H., C. Meghir, and L. Pistaferri (2009), "Wage risk and employment risk over the life cycle." Working Paper, NBER. [473]

MaCurdy, T. (1981), "An empirical model of labor supply in a life-cycle setting." *The Journal of Political Economy*, 89 (6), 1059–1085. [473]

McDaniel, C. (2007), "Average tax rates on consumption, investment, labor and capital in the OECD 1950–2003." Report, University of Arizona. [481]

Menzio, G., I. Telyuokva, and L. Visschers (2010), "Directed search over the lifecycle." Report, University of Pennsylvania. [473, 489]

Pijoan-Mas, J. (2006), "Precautionary savings or working longer hours?" *Review of Economic Dynamics*, 9 (2), 326–352. [473]

Prescott, E., R. Rogerson, and J. Wallenius (2009), "Lifetime aggregate labor supply with endogenous workweek length." *Review of Economic Dynamics*, 12 (1), 23–36. [489]

Primiceri, G. and T. Van Rens (2009), "Heterogeneous life-cycle profiles, income risk and consumption inequality." *Journal of Monetary Economics*, 56 (1), 20–39. [478]

Rogerson, R. and J. Wallenius (2009), "Micro and macro elasticities in a life cycle model with taxes." *Journal of Economic Theory*, 144 (6), 2277–2292. [489]

Slesnick, D. (2005), *Consumption and Social Welfare*. Cambridge Books, Minneapolis, MN. [478]

Slesnick, D. and A. Ulker (2004), "Inequality and the life-cycle: Age, cohort effects and consumption." Report, University of Texas–Austin. [478, 513]

Storesletten, K., C. Telmer, and A. Yaron (2001), "How important are idiosyncratic shocks? Evidence from labor supply." *American Economic Review*, 91 (2), 413–417. [472, 485]

Storesletten, K., C. Telmer, and A. Yaron (2004), "Consumption and risk sharing over the life cycle." *Journal of Monetary Economics*, 51 (3), 609–633. [471, 481]

Yang, Y., S. Schulhofer-Wohl, W. Fu, and K. Land (2008), "The intrinsic estimator for age-period-cohort analysis: What it is and how to use it." *American Journal of Sociology*, 113 (6), 1697–1736. [478, 484]

Zhang, H., A. Conn, and K. Scheinberg (2010), "A derivative-free algorithm for the least-squares minimization." *SIAM Journal on Optimization*, 20 (6), 35–55. [518]

Submitted October, 2011. Final version accepted February, 2012.