University of Toronto
Department of Economics

Working Paper 443

Towards a Micro-Founded Theory of Aggregate Labor Supply

By Andres Erosa, Luisa Fuster and Gueorgui Kambourov

November 23, 2011

Towards a Micro-Founded Theory of Aggregate Labor Supply*

Andrés Erosa[†]
IMDEA Social Sciences Institute

Luisa Fuster[‡]
IMDEA Social Sciences Institute

Gueorgui Kambourov[§]
University of Toronto and NBER

Abstract

We build a heterogeneous life-cycle model which captures a large number of salient features of individual labor supply, by education, over the life cycle. The model provides an aggregation theory of individual labor supply, firmly grounded on micro evidence, and is used to study the aggregate labor supply responses to changes in the economic environment. We find that the aggregate labor supply elasticity to a transitory wage shock is 1.27, with the extensive margin accounting for 54% of the response. Furthermore, we also simulate the 1987 tax holiday in Iceland − a quasi-natural experiment − and find that the aggregate labor supply responses in the model are similar to those actually observed in Iceland.

JEL Classification: D9, E2, E13, E62, J22.

Keywords: Aggregate labor supply, intensive margin, extensive margin, heterogeneous agents, life cycle

*This version: November 23, 2011. This paper was previously circulated under the title "The Heterogeneity and Dynamics of Individuals Labor Supply over the Life Cycle: Facts and Theory." We thank seminar participants at Alberta, the Atlanta Fed, Carlos III, the Chicago Fed, Concordia, GREQAM, IIES Stockholm, Iowa, the Minneapolis Fed, Montreal, MOVE Barcelona, Pennsylvania, Queen's, Rochester, Simon Fraser, SUNY Buffalo, Toronto, UAB, UBC, UWO, Zurich, the 2008 Yale-Cowles Summer Conference, the 2008 CEA Conference, the 2009 SED Conference, the III Dynamic General Equilibrium Macroeconomics Workshop, the 2008 WCU-Seoul Conference on Recent Development in Macroeconomics, the 2009 NBER Summer Workshop, the 2009 Minnesota Macroeconomic Workshop, the 2010 meetings of the European Economic Association, the 2010 Rimini Conference in Economics and Finance, the 2010 Conference on Labor Markets in Santa Barbara, the 2010 Canadian Macroeconomics Study Group meeting, the 2011 AEA meetings, and the 2011 Vienna Macroeconomics Workshop for useful comments and discussions. This research was supported by a Marie Curie International Reintegration Grant within the 7th European Community Framework Programme and the Social Sciences and Humanities Research Council of Canada. We thank Hui Xiong for excellent research assistance.

[†]IMDEA Social Sciences Institute, Calle Isaac Newton 2, Tres Cantos, 28760 Madrid, Spain. E-mail: andres.erosa@imdea.org

[‡]IMDEA Social Sciences Institute, Calle Isaac Newton 2, Tres Cantos, 28760 Madrid, Spain. E-mail: luisa.fuster@imdea.org

[§]Department of Economics, University of Toronto, 150 St. George St., Toronto, ON, M5S 3G7 Canada. E-mail: g.kambourov@utoronto.ca.

1 Introduction

How responsive is aggregate labor supply to changes in the economic environment? The answer to this question has wide-ranging implications for understanding the effects of numerous phenomena such as business cycles and various government policies.[1] Aggregate labor supply is ultimately the sum of all individuals' labor supply decisions. The empirical evidence indicates that aggregate labor supply responses are determined by individual responses along both the intensive and the extensive margins.[2] Further, economic theory implies that labor supply responses along the intensive and extensive margins are distinct objects. The intensive margin responses are mainly driven by the intertemporal substitution of labor (the Frisch elasticity of labor supply). The extensive margin responses, on the other hand, are determined by the distribution of reservation wages and the mass of agents who are close to being indifferent between working or not. Chang and Kim (2006) build on the insights from the model of indivisible labor in Rogerson (1988), introduce heterogeneity, and show that the slope of the aggregate labor supply schedule is determined by the distribution of reservation wages rather than by the willingness to substitute leisure intertemporally, establishing that when the extensive margin is operative heterogeneity and aggregation play a crucial role in determining aggregate labor supply responses.

Individuals differ along a large number of dimensions and, depending on their characteristics, will respond differently to aggregate and idiosyncratic shocks, such as a change in taxes or a labor productivity shock − the young would respond differently from the old, the non-college from the college, those with a small amount of assets from those with a large amount of assets. Therefore, in order to have a theory of aggregate labor supply behavior, which can then be used with some confidence in analyzing the effects of various government policies and macroeconomic shocks, the starting point needs to be a quantitative theory which is rich in heterogeneity and is consistent with individuals' labor supply behavior along many dimensions. Only then, having been disciplined with micro-level facts, can the model become a useful tool for explicitly aggregating individuals' decisions into an aggregate labor supply response. This is the goal of this paper.

We proceed with our analysis as follows. We start by presenting a rich set of facts about the labor supply decisions of male workers in the US, by education, over their life cycle using data from the Panel Study of Income Dynamics (PSID) as well as the Survey of Income and

[1] See Keane (2010) and Keane and Rogerson (2011) for a recent survey of the literature.

[2] See Cooley, ed (1995) for evidence on the adjustment in labor supply along both margins over the business cycle and Blundell et al. (2011) for recent evidence on the importance of both margins over time in the US, the UK, and France.

Program Participation (SIPP).[3] Next, we develop a neo-classical model of the labor market with heterogeneous agents who make labor supply decisions both along the extensive margin (whether to work or not) and the intensive margin (how much to work). The key feature of our theory for delivering periods of non-participation is the nonlinear mapping between hours of work and earnings, which is convex at low hours of work. This mapping is obtained as the competitive equilibrium outcome of an economy with a production technology in which hours of work and number of workers are imperfect substitutes; e.g., see Hornstein and Prescott (1993) and Osuna and Ríos-Rull (2003). The nonlinearity in earnings is disciplined with the evidence in Aaronson and French (2004) who use the exogenous variation in the US social security rules and find that a 50% decline in hours worked decreases hourly wages by 20%.

We obtain heterogeneity in the model by introducing life cycle, education, and incomplete markets. The theory models life-cycle behavior in order to relate better the model predictions to the data – the various labor supply facts presented in the empirical section exhibit strong life-cycle patterns. Furthermore, as will become evident from our analysis, the labor supply responses of heterogeneous agents, especially along the extensive margin, interact in an important way with the incomplete markets at various stages over the life cycle. Chang and Kim (2006), Domeij and Flodén (2006), and Pijoan-Mas (2006) study labor supply decisions in a framework with heterogeneous agents and incomplete markets but, differently from our paper, they do not model life-cycle behavior. The model in French (2005) which features a life cycle, incomplete markets, and nonlinear earnings is the closest to our framework. However, we address very different issues – while French (2005) focuses mainly on retirement behavior, we study the labor supply responses to transitory and permanent shocks over the whole life cycle and model the extensive margin at a 4 month rather than annual period.[4] Relative to Rogerson and Wallenius (2009), who build a life-cycle (complete markets) model with an operative extensive margin, our contribution is to build a rich in heterogeneity theory of aggregation which is disciplined with micro data. We will show that the aggregate response to temporary unexpected shocks in our incomplete markets economy is significantly lower

[3] We have also documented similar facts for females, but in this paper we focus only on the aggregate labor supply of males. While it would also be important to model female labor supply, this is left for future research. First, there are dramatic changes in labor supply across different cohorts of women, and modeling them will require a non-stationary environment and taking a stand on the causes for these changes. Second, even in a stationary environment, it is a daunting computational task to model a two-earner household in an incomplete-markets model with a sub-annual period. Nevertheless, we conjecture that the insights from our current analysis of male labor supply would be also valuable for understanding aggregate female labor supply responses.

[4] Low (2005) studies individual labor supply in a life-cycle model with incomplete markets, but, differently from our approach, his model does not feature a 4-month model period, nonlinear earnings and an active extensive margin, and does not study aggregate labor supply responses.

than in the Rogerson and Wallenius (2009) economy and is consistent with evidence from the quasi-natural experiment provided by the 1987 tax holiday in Iceland (see below for further discussion).

The calibration of the model economy involves three key tasks. First, we pin down the parameter determining how hours of work affect labor productivity in the model economy using estimates from Aaronson and French (2004). In a second step, the age profile and shock process on labor productivity are estimated following an indirect inference approach that explicitly controls for the selection problems that make the calibration of these parameters difficult. The third task is to take a stand on measurement error in hours in the PSID data. We propose a novel approach to estimate measurement error that consists in comparing both in the model and in the data the variance of transitory wages in two alternative specifications of the wage process. The first specification estimates the process for observed wages while the second specification estimates the process for wages net of the effect of hours of work on wages. Identification comes from the fact that measurement error in hours has a different effect on the variance of transitory wages in the two specifications of the estimation.

We calibrate alternative economies that differ in the preference parameter σ determining the intertemporal elasticity of substitution of leisure (i.e.s.) given by $1/\sigma$. We find that when the i.e.s. varies in the tight range of 0.45 to 0.55 (σ between 1.8 and 2.2) the model captures well the life-cycle patterns in hours worked for college and non-college individuals in the US, even though these patters were not explicitly targeted by the calibration. The baseline economy (σ equal to 2) accounts for the low co-movement of hours and wages early in the life cycle, when wages are rising rapidly but hours are relatively flat. Incomplete markets are crucial for this result. While individuals face an increasing age profile of wages, they work long hours when young because they need to build a buffer stock of savings to self-insure against income risk. By age 50 their stock of assets is sufficiently large that individuals can afford to take a quadrimester off work when they receive a low realization of the temporary wage shock. This accounts for the pronounced decline in annual working hours late in the life cycle. The model is also consistent with most of the salient features about labor supply in the data such as heterogeneity in labor supply, persistence in annual hours, and covariance and correlation of changes in hours and wages. We find that measurement error in hours is quantitatively important and varies with the education and age of individuals (over the life cycle the variance of measurement error in log hours is on average slightly less than 0.03 for non-college individuals and about 0.02 for college individuals).

Given that the theory accounts well for the micro facts, we use the model economy to study aggregate labor supply responses. We find that a one period wage change in the baseline economy implies an aggregate labor supply (Frisch) elasticity of 1.27, which is more than twice as big as the Frisch (theoretical) elasticity embedded in the calibration of the model economy (0.61). The large labor supply response along the extensive margin explains why the Frisch elasticity is much bigger than the theoretical elasticity. Restricting attention to labor supply changes only along the intensive margin decreases the Frisch elasticity from 1.27 to 0.58. Hence, the extensive margin accounts for about 54% of the aggregate labor supply response to a temporary wage change. This finding is consistent with the evidence provided in Kimmel and Kniesner (1998) who estimate in SIPP data an elasticity of labor supply of 1.25, with the extensive margin accounting for 70% of the total elasticity. To evaluate the compensated elasticity to a permanent wage change, we simulate an increase in the labor income tax. Tax proceeds are assumed to be rebated with a lump sum transfer to working-age individuals. We find that the elasticities for both the intensive and extensive margins are reduced by a half relative to the case of a temporary wage change. Now the labor supply elasticity is 0.65 and the employment elasticity is 0.35, which should be compared to the elasticities of 1.27 and 0.69 to a temporary wage change. Hence, not surprisingly, individuals respond more strongly to a temporary wage change than to a permanent compensated-wage change.

To put our findings in perspective, we find it convenient to conclude the discussion by answering the following three questions. First, are the aggregate labor supply responses in our calibrated model economy consistent with the quasi-experimental evidence surveyed in Chetty et al. (2011)? The mean value of the Hicks elasticity of aggregate hours across the micro studies reviewed in Chetty et al. (2011) is 0.76, with substantial variation in the estimated elasticities across studies.[5] Consistently with this evidence, our baseline economy predicts a Hicks elasticity of aggregate hours of work of 0.65. In order to evaluate the predictions of our theory of labor supply, we simulate the tax holiday that took place in Iceland in 1987. The Icelandic tax holiday is ideally suited for identifying intertemporal labor supply responses because it induced an unanticipated temporary wage variation during the year 1987. In 1987, Iceland moved from a system under which taxes were paid on the previous year's income to a pay-as-you-earn system. The transition to the new tax system implied that income during 1987 was never taxed since the tax base in 1987 was income earned in 1986 and the tax base in 1988 was income earned in 1988. The average tax rate was 14.5% in 1986, 0% in 1987, and 8.0% in 1988. We simulate in our baseline economy a one year (three model periods)

[5]Since there are wide confidence intervals associated with each of the point estimates as well as methodological disputes about the validity of some of the studies, Chetty et al. (2011) argue that the estimates should be treated as rough values meant to gauge the order of magnitudes.

reduction in the tax rate of 14.5 percentage points that is followed by a permanent decrease of 6.5 percentage points in the average tax rate (relative to the initial tax rate of 27% in the baseline economy). We find that the aggregate elasticity of labor supply implied by the Icelandic tax holiday experiment is 0.68, with extensive and intensive margin elasticities of 0.33 and 0.35, respectively. Remarkably, these elasticity results are not far from the ones estimated by Bianchi et al. (2001) in the Icelandic micro data: for male workers, they find an employment elasticity of 0.58 and an intensive margin elasticity of 0.26. Our theory is also consistent with the evidence on the intertemporal substitution elasticities in labor supply late in the life cycle documented in the cross-country studies in Gruber and Wise, eds (1999). In Erosa et al. (2012) we extend our baseline economy to model in detail the variation in the social security, disability insurance, and taxation institutions across European countries and the United States. We find that the extended baseline model economy accounts well for the observed cross-country differences in labor supply late in the life cycle, indicating that the Frisch elasticity of labor supply in our model economy is plausible.

Second, does heterogeneity play an important role in our results? This question can be answered by comparing the aggregate labor supply responses in our model to those in Rogerson and Wallenius (2009) (RW), who evaluate a life-cycle model with nonlinear earnings and complete markets. While labor supply responses to permanent tax changes are similar across both models, the RW model predicts much higher responses to temporary tax changes. Chetty et al. (2011) simulate the Icelandic tax experiment on the RW model and find that it implies an implausibly high change in employment. On the other hand, our model's predictions for the labor supply responses are consistent with the evidence from the Icelandic tax holiday. Our baseline economy with heterogeneous agents and incomplete markets has a smaller fraction of agents that are close to being indifferent between working or staying out of the labor force than in the Rogerson and Wallenius (2009) representative agent model.[6] We thus conclude that heterogeneity is key for understanding aggregate labor supply responses to temporary shocks (Frisch elasticity) but not to permanent tax changes (Hicks elasticity).

Third, is the modeling of nonlinear earnings important for our findings? By modeling nonlinear earnings labor supply also responds along the extensive margin, and the aggregate elasticity becomes substantially larger than the theoretical elasticity embedded in the calibration of the preference parameters. In fact, when we simulate a version of our model

[6] In the language of Ljungqvist and Sargent (2011), and relative to the RW model, our economy with heterogeneity within cohorts has more individuals that are at a corner in their labor supply decisions at the participation margin (i.e., far from being indifferent between working or not). In addition, young individuals, who are building precautionary savings, are less responsive at the intensive margin to temporary shocks.

economy with linear wages the aggregate labor supply response is close to the one implied by the theoretical Frisch elasticity. Hence, the extensive margin is an important factor for generating large aggregate labor supply responses. It is important to emphasize this point since several recent papers on labor supply imply a smaller aggregate Frisch elasticity than the Frisch elasticity embedded in the individuals' preferences.[7]

The paper is structured as follows. Section 2 presents empirical facts on labor supply using data from the PSID and the SIPP. Section 3 develops a life-cycle theory of individual labor supply with heterogeneous agents. The calibration of the model economy is discussed in Section 4. Section 5 discusses the performance of the baseline economy in accounting for the documented facts on labor supply. Section 6 studies aggregate labor supply responses and evaluates the effect of temporary and permanent wage changes on aggregate labor supply. Section 7 concludes.

2 Empirical findings

2.1 The data

We use the Michigan Panel Study of Income Dynamics (PSID) for the period 1968-1997 in order to compute all annual statistics.[8] The sample is restricted to males between the ages of 18 and 65. We do not place other restrictions on the sample. In particular, note that we do not restrict to heads of households − we use the information on annual hours worked provided by the PSID for those males who are listed as "wives" as well as the information on annual hours worked, whenever available in the individual files, on males who are dependents. This allows us to provide a more representative overview of the facts on labor supply as compared to the related literature which has mainly focused on male workers with strong labor market

[7]Imai and Keane (2004) emphasize the importance of human capital accumulation while Domeij and Flodén (2006) and Pijoan-Mas (2006) study the effect of borrowing constraints on understanding the individual labor supply responses. While Imai and Keane (2004) estimate a large Frisch elasticity of labor supply, the aggregate labor supply response to a temporary wage shock is four times smaller. Intuitively, when the returns to human capital are an important part of the return to work, temporary wage shocks have a smaller effect on the incentives to work, and labor supply responses are small. Similarly, but through a different economic mechanism, the labor supply response of liquidity constrained individuals will be smaller, or even of the opposite sign, than what is predicted by an analysis that ignores such constraints, as emphasized by Domeij and Flodén (2006) and Pijoan-Mas (2006). Note, again, that this reasoning also implies that labor supply should not be very responsive to aggregate temporary wage shocks − a result which we confirm in an experiment in which we simulate a linear earnings version of our benchmark economy.

[8]We have performed a similar empirical analysis using the Survey of Income and Program Participation (SIPP). The annual statistics obtained on the SIPP data are largely consistent with those obtained on the PSID data. These are available from the authors upon request.

attachment.[9]

The analysis is focused on the labor supply of men. A cohort is defined to consist of all individuals who turn 18 years old in a given year − for example, the 1967 cohort consists of all individuals who turn 18 years old in 1967. Since the PSID is a relatively small dataset, we grouped our sample into age and cohort groups. By age, individuals are grouped into 12 age groups each consisting of four ages − for example, the age-18 group on the graphs includes individuals between the age of 18 and 21, while the age-22 group includes all individuals between the ages of 22 and 25. We have 17 cohort groups each consisting of three cohorts − for instance, the 1976 cohort group includes cohorts 1976, 1977, and 1978 while the 1985 cohort includes cohorts 1985, 1986, and 1987. We drop all cohorts smaller than 1940 and all cohorts greater than 1990.[10] We use PSID sample weights in the analysis.

Next we proceed with the empirical analysis and document a rich set of facts regarding the labor supply of men over the life cycle. The patterns that we see in the data will be motivating the main features which will be introduced in the model. The most important patterns are as follows: i) we see a very pronounced life-cycle pattern in the labor supply behavior of men. We see a life-cycle trend in the mean annual hours worked, the participation rate, and the dispersion of annual hours; ii) there is a substantial dispersion of annual hours worked at every point in the life cycle; iii) for most individuals, and for most ages during the life cycle, annual hours are quite persistent; and iv) the labor supply behavior of high school and college graduates is different enough to warrant a separate analysis for each of these groups.[11]

2.2 Facts on the life-cycle labor supply of men

2.2.1 Average annual hours over the life cycle

Figure 1 shows that mean annual hours worked clearly exhibit an inverted U-shape over the life cycle − they increase early in life until the late 20s, stay constant after that until the late 40s, and decline monotonically after the age of 50. The second panel shows that college and non-college graduates have different life-cycle profiles − college graduates initially work less (while studying) while working more after the age of 26. In addition, the mean annual hours of high-school workers start declining earlier, at the age of 50.

[9] See for example Storesletten et al. (2001), Heathcote et al. (2010), Kaplan (2011).

[10] When we conduct the analysis by education groups, our last cohort is 1985 in order to be able to classify individuals as either high school or college.

[11] We consider an individual to be high school if he or she has at most 12 years of education while those with 14 years of education or more are considered to be college graduates. A sensitivity analysis with respect to the education cut-off separating high-school and college graduates indicates that the current partition is a sensible one.

Figures A-1 and A-2 illustrate the intensive and extensive margins at the annual level of labor supply of men over the life cycle. Between ages 30 to 46 working hours are quite constant and average annual hours are about 2,200 for non-college and 2,300 for college graduates. The extensive margin at the annual level matters early in life until the age of 26, but is especially quantitatively important late in life after the age of 50. Furthermore, it is interesting to point out that the participation rate of those with high-school starts declining in the late 40s while the participation rate of those with college start declining significantly only in the late 50s.

2.2.2 Dispersion of annual hours over the life cycle

Figure 2 displays the dispersion of annual hours over the life cycle as measured by the coefficient of variation of annual hours. This figure illustrates three facts of particular importance. First, the dispersion in annual hours is U-shaped − it is high early in the life cycle until the age of 26, then declines and is constant until the late 40s, and increases substantially after the age of 50. Second, the degree of dispersion is quite substantial as the coefficient of variation is between 0.30 and 0.40 during the life cycle, except close to retirement when it increases above 1. Third, even though the dispersion of hours over the life cycle has the same shape for both groups of college and non-college individuals, the coefficient of variation in hours is higher for the non-college group for all ages after 22.

2.2.3 Persistence in annual hours worked

In this section, we investigate the extent to which annual hours worked are persistent over the individual's life. For that purpose each year we divide individuals into four groups: 1 − those with annual hours less than 100; 2 − those with annual hours between 100 and 1500; 3 − those with annual hours between 1500 and 2800; and 4 − those with annual hours greater than 2800.[12] We then construct transition matrices where cell ij indicates the fraction of all individuals in cell i in year t who moved to cell j in year $t+1$. We document the facts for all men, as well as for high school graduates and college graduates.

Table A-3 presents the transition matrix and the relative size of each group of men in three age groups: young workers between the ages of 18 and 29, middle-aged workers between the ages of 30 and 54, and old workers between the ages of 55 and 65. We found it useful to present graphically some of these results. In particular, Figure 3 graphs the relative size of

[12] The cut-offs were chosen in order to capture four broad patterns of labor market behavior − no labor market participation (group 1), part-time labor supply (group 2), full-time labor supply (group 3), and very high labor supply (group 4). Slight changes in these cut-offs do not significantly change the main patterns documented here.

each of the four groups as well as the fraction of workers who stay in each of these groups in two consecutive years (i.e. the diagonal elements from the transition matrix). Note that this graphical representation makes it easy to consider 12 age groups rather than the 3 age groups considered in Table A-3.

Three important findings are worth pointing out. First, the group of full-time workers with annual hours between 1500 and 2800 is by far the largest, with the exception of the first and very last years of the life cycle, and exhibiting very high persistence in annual hours worked − over 70% of men are in this labor supply group and more than 80% of those who are in this group in year t remain in it in year $t+1$. Table A-3 further shows that, between the ages of 30 and 54, most of those who move out of this group move temporarily into the group with large labor supply and work more than 2800 hours. That indicates that for the most part of the life cycle, especially between the ages of 30 and 50, annual hours worked are quite persistent for most men. Second, the fraction of men who work less than 100 hours is quite small throughout the life cycle, but starts increasing gradually after the age of 46. Furthermore, with age, this group becomes an absorbing state − after the age of 46, more than 80% of men who are in this group in year t will be there in year $t+1$. Furthermore, as Table A-3 shows, those who move out of it later in life, move temporarily into the part-time labor supply group. Third, the other two groups − those working between 100 and 1500 hours and those working more than 2800 hours − do exhibit a life-cycle pattern but are relatively small. In addition, each of these two groups seem to represent a temporary state in one's labor market history since the probability of remaining there is not very high.

These broad patterns are observed also for each of the two education groups − high-school and college men (see Tables A-4 and A-5). After the age of 30, the group of full-time workers with annual hours between 1500 and 2800 (i) is bigger for the college men than the high-school men, (ii) starts declining earlier for high-school men than college men, and (iii) is more persistent for college than for high-school men.

2.2.4 Lifetime labor supply

Using the fact that the PSID is a long panel we show that there is no association between lifetime labor supply and average labor productivity (wages) over the life cycle across individuals of the same education group. To this end, for each cohort and education group we divide individuals into high and low productivity types. First, we compute each individual's mean wage over the age of 30 to 45 and classify them into high and low types depending on whether their mean wages are above or below the median wage in their cohort-education category. We then compute mean hours worked for high and low types and find that there are virtually no differences in labor supply.[13]

The dispersion in lifetime labor supply is another useful statistic which is closely related to the persistence in an individual's labor supply over time. Due to the nature of the PSID dataset, we do not observe individuals throughout all their life − some of them have already been in the labor market for some time when the survey starts in 1968 while those who enter the labor market in 1968 at the age of 18 are only in their 40s in 1997. Nevertheless, we can learn a lot even if we follow individuals for shorter periods. We choose to follow individuals for periods of 10 years at different stages in their life-cycle: ages 26-35, 36-45, 46-55, and 56-64. We drop all individuals who have a missing observation during the relevant ten years and sum the hours worked for each individual during the whole ten years. Then we compute the dispersion in this cumulative measure of hours worked. Considering two extreme examples is useful for illustrating how to interpret the results. Consider a particular group, e.g. the group between the ages of 36 and 45, and suppose that all individuals work the same number of hours throughout the whole period as at the beginning at the age of 36. In that case, the coefficient of variation of the cumulative hours worked throughout the whole period would be the same as the coefficient of variation (cross-sectionally) at the age of 36 (or any other age in the period). Alternative, suppose that individual hours fluctuate a lot over the period and those who work a lot in one year work very little the year after that. In that case, workers would end up working quite similar cumulative hours over the period, and the coefficient of variation of the cumulative hours worked throughout the whole period would be quite small and substantially lower than the coefficient of variation (cross-sectionally) at the age of 36 (or any other age in the period).

Table 1 reports the coefficient of variation of the cumulative hours worked for the four age groups defined above. The results indicate that hours are quite persistent, especially over the ages of 26 and 55. This analysis provides us with two important findings. First, the dispersion in cumulative hours is quite substantial, indicating that individuals tend to be quite persistent in their labor supply behavior. This is consistent with the mobility matrices discussed in section 2.2.3. Second, the dispersion of cumulative hours is smaller than the cross-sectional dispersion at any age in the 26-64 interval. This implies that workers do sometimes change their hours worked. This is also consistent with the mobility matrices discussed in section 2.2.3 since − as seen in the middle panel of Table A-3 for those between the ages of

[13] Focusing on the age group 30-45 and non-college individuals, the average hours worked across all cohorts is 2143 for type 1 individuals and 2166 for the type 2. For individuals with college education, average hours are 2269 and 2271 for type 1 and type 2, respectively.

30 and 54 – the diagonal elements of the mobility matrices are not zero, and we do observe workers who switch across the hours categories.

2.3 The Survey of Income and Program Participation (SIPP)

We also use data from the Survey of Income and Program Participation (SIPP). The SIPP interviews individuals three times a year (rather than once a year as in the PSID) and allows us to compute other labor market statistics of interest at a lower frequency, such as a quadrimester (a 4-month period). We use the 1990 SIPP Panel. Figure A-3 shows the distribution of hours within a quadrimester, for six age-education groups. The distribution of quadrimesterly hours is bimodal, with a peak at zero hours and another one at 600 hours. Further, there are virtually no individuals working between zero and 500 hours in a quadrimester. Figure A-4 graphs the fraction of individuals working three quadrimesters in a year. There is a sizable fraction of individuals that do not work all year round – more than 20% of the non-college and more than 10% of the college.

3 The Model

We develop a life-cycle theory of the labor supply of individuals. The model abstracts from the labor supply decisions of women and analyzes only males. We consider a small open economy facing a fixed interest rate and follow Hornstein and Prescott (1993) in modeling a production technology that gives rise to a competitive equilibrium with nonlinear earnings. The model economy is closely related to French (2005).

3.1 Population, preferences, and endowments

The economy is populated by overlapping generations of individuals who start their lives at age 25, face uncertain lifetimes, and live, at most, J periods. They differ in terms of their education (college versus non-college) and labor productivity. The date-t utility function takes the form

$$u_t = u(c_t, l_t) = \ln c_t + \varphi \frac{l_t^{1-\sigma}}{1-\sigma}, \tag{1}$$

where c_t is consumption and l_t denotes leisure. The utility function is consistent with balanced growth – this assumption allows the theory to be consistent with the fact, discussed in section 2.2.4, that there are large permanent differences in labor productivities across individuals (heterogeneity in fixed effects) but not in their lifetime labor supply. Note that by modeling the utility of leisure (rather than the disutility of labor), the theory allows for an active extensive margin. In particular, the specification $u(c_t, h_t) = \ln c_t - \varphi \frac{h_t^{1-\sigma}}{1-\sigma}$ does not deliver an active extensive margin, and, moreover, it often implies that individuals work 100% of their available time.

Individuals maximize lifetime expected utility

$$E_t \sum_{j=t}^{J} \beta^{j-t} u(c_j, l_j), \tag{2}$$

where E_t denotes expectations at date-t. Individuals face mortality shocks each period and uncertainty regarding their labor productivity z up to age 65 when labor productivity is zero (mandatory retirement). An individual's time endowment in each period is one. The amount of time that can be allocated to work is $h_j = 1 - l_j$. The college decision is exogenous, and the education type of an individual determines the stochastic processes driving the mortality and labor productivity shocks.

3.2 Technology

There are a large number of plants, and each plant is a collection of jobs. We assume that plants can operate jobs at zero costs. The production function of a job is given by

$$f(K, h, z) = h^{\varepsilon} K^{1-\theta} z^{\theta}, \quad \text{with } \theta \leq \varepsilon \leq 1 \tag{3}$$

where h denotes the workweek, K is the amount of capital for the job, and z is effective labor in the job (which is given by the worker productivity). Note that, for a fixed workweek, the job technology exhibits constant returns to scale in capital and effective labor. Moreover, as discussed in Osuna and Ríos-Rull (2003), when $\varepsilon = \theta$ the job technology reduces to the standard Cobb-Douglas technology where total hours of effective labor is what matters. When $\varepsilon > \theta$ the hours and effective labor are imperfect substitutes and the composition between these two inputs matters. When $\varepsilon = 1$ the technology is linear in hours and corresponds to the case where workers are not subject to fatigue. While the production function of a job features increasing returns in the three inputs, Hornstein and Prescott (1993) show that the aggregate technology set is convex. Intuitively, the economy's output doubles when the productive resources in the economy double (labor force and capital).

3.3 The plant's problem

The plant takes as given the earnings schedule $\tilde{w}(h, z)$ and the interest rate r. For each job, the plant chooses hours of work h, capital K, and effective labor z. In equilibrium, capital is paid its marginal product, and the competition for labor implies that workers will be the residual claimants on the output which remains after capital has been paid and that profits will be zero. Moreover, the earnings schedule is a nonlinear function of the workweek h and a linear function of effective labor z. To show this point, consider a job hiring a worker, with z units of effective labor, for h hours. The optimal amount of capital K solves

$$\pi = \max_{K}\{h^{\varepsilon} K^{1-\theta} z^{\theta} - K(r+\delta) - \tilde{w}(h,z)\}. \tag{4}$$

Then, the solution to this problem implies that

$$\frac{K}{z} = k^{*}(h,r) = \left[\frac{(1-\theta)h^{\varepsilon}}{r+\delta}\right]^{1/\theta}. \tag{5}$$

Next, notice that a job is open only if profits are non-negative. Free entry, and the fact that jobs can be opened at zero cost, imply that in equilibrium plants will make zero profits from each job. Hence, competition for workers implies that the wage bill $\tilde{w}(h,z)$ is determined from

$$\pi = h^{\varepsilon}\left[zk^{*}(h,r)\right]^{1-\theta} z^{\theta} - zk^{*}(h,r)(r+\delta) - \tilde{w}(h,z) = 0, \tag{6}$$

which gives

$$\begin{aligned} \tilde{w}(h,z) &= zw(h), \text{ where} && (7)\\ w(h) &\equiv (r+\delta)\frac{\theta}{1-\theta}\left[\frac{(1-\theta)}{r+\delta}\right]^{1/\theta} h^{\frac{\varepsilon}{\theta}} = \Theta h^{\frac{\varepsilon}{\theta}}. && (8) \end{aligned}$$

It follows that the earnings schedule $\tilde{w}(h,z)$ is linear in effective labor z and nonlinear in hours of work h. When $\varepsilon = \theta$ earnings are also linear in h. When $\varepsilon > \theta$ earnings increase more than proportionally with h and, hence, the hourly wage rate also increases with h. In this case, households would be better off by selling employment lotteries to firms (Hornstein and Prescott (1993)). However, we rule out this possibility by assuming that households cannot commit to work when the realization of the employment lottery implies that they should work.

Notice that the log hourly wage of an individual with labor productivity z and working h hours is

$$\ln w_h(h) \equiv \ln\frac{zw(h)}{h} = \ln\Theta + \left(\frac{\varepsilon}{\theta} - 1\right)\ln h. \tag{9}$$

This is the functional form typically estimated in the empirical literature on nonlinear wages (see the discussion in Aaronson and French (2009)). The theory presented above provides a theoretical rationale for the functional form used in these empirical studies. Moreover, this is relevant since our calibration strategy will use the estimates from Aaronson and French (2004) to pin down $\frac{\varepsilon}{\theta}$.

3.4 Government, annuity, and credit market

The government taxes consumption, capital income, and labor income. The tax revenue is used to finance government expenditures. Individuals can insure mortality risk in fair annuity markets. Denoting by R the gross interest rate net of capital income taxes τ_k, the gross interest rate faced by an individual j years old with education e is given by

$$R_j^e = 1 + \left(\frac{1+r}{\pi_j^e} - 1\right)(1-\tau_k), \tag{10}$$

where π_j^e is the conditional probability that an age $j-1$ individual with education e survives to age j. We assume that individuals cannot borrow.

Social security. The government also administers a pay-as-you-go social security system. To finance pensions for retired individuals, the government uses a payroll tax τ_{ss}. Individuals retire at age 65. Social security benefits depend on the average earnings made by individuals over the 35 highest years of earnings. Denoting this average earnings by $\overline{w}$, social security benefits can be expressed as $b_s(\overline{w})$.

Social security benefits are a function of the Average Indexed Monthly Earnings (AIME) over the 35 highest earnings years. Given that the model period is a quadrimester, for computational simplicity we compute average quadrimesterly earnings over the 35×3 highest quadrimesterly earnings as follows

$$\overline{w}_{j+1} = \overline{w}_j + zw(h_j)/(35 \times 3) \text{ for } j \leq 35 \times 3, \tag{11}$$

$$\overline{w}_{j+1} = \overline{w}_j + \max\{0, (\min\{zw(h_j), \widehat{y}\} - \overline{w}_j)/(35 \times 3)\} \text{ for } j > 35 \times 3, \tag{12}$$

where $\widehat{y}$ is the maximum taxable earnings by the social security administration, which is set at 2.47 the average earnings in the economy. We express (11)-(12) in the following compact way:

$$\overline{w}' = \Gamma_{ss}(\overline{w}, zw(h)). \tag{13}$$

At retirement, the Social Security Administration computes the Primary Insurance Amount (PIA) which is the sum of three portions of the Average Index Monthly Earnings (AIME). The bend points in the PIA formula are 0.2 and 1.24 of the average earnings in the economy when individuals file for social security ($\overline{W}$).[14] The social security benefit is given by the

[14] $\overline{W}$ is the average earnings in the economy in the year when the individual becomes 62 years old.

following formula:

$$b_s(\overline{w}) = \begin{cases} 0.90 \times \overline{w} \text{ for } \overline{w} < 0.2\overline{W}, \\ 0.90 \times 0.2\overline{W} + 0.33 \times (\overline{w} - 0.2\overline{W}) \text{ for } \overline{w} \in (0.2\overline{W}, 1.24\overline{W}], \\ 0.90 \times 0.2\overline{W} + 0.33 \times (1.24\overline{W} - 0.2\overline{W}) + 0.15 \times (\overline{w} - 1.24\overline{W}) \text{ for } \overline{w} > 1.24\overline{W}. \end{cases} \tag{14}$$

3.5 The individual's problem

We use recursive language to describe the problem of an individual. To simplify the notation, we abstract from the fact that the education type of an individual determines his earnings and mortality processes. Then, the state of an individual is given by his age j, assets a, average lifetime earnings $\overline{w}$, and earnings shock z.

Since individuals live for at most J periods, we set $V_{J+1}(\cdot) = 0$. Individuals retire exogenously at age 65. The value of a retired person is given by

$$\begin{aligned} V_j(a, b_s) &= \max_{\{c,a'\}} \{u(c,1) + \beta\pi_{j+1} E[V_{j+1}(a', b_s')]\} \qquad (15) \\ \text{s. t.} & \\ a' &= b_s + R_j a - c(1+\tau_c), \\ a' &\geq 0. \end{aligned}$$

The value of a person who has not retired is

$$\begin{aligned} V_j(a, \overline{w}, z) &= \max_{\{c,h,a'\}} \{u(c,l) + \beta\pi_{j+1} E[V_{j+1}(a', \overline{w}', z')]\} \qquad (16) \\ \text{s. t.} & \\ a' &= (1 - \tau_{ss} - \tau_h + 0.5\tau_{ss}\tau_h)\min\{zw(h), \widehat{y}\} + (1-\tau_h)\max\{zw(h) - \widehat{y}, 0\} \\ &\quad + R_j a - c\,(1+\tau^c), \\ a' &\geq 0, \\ \overline{w}' &= \Gamma_{ss}(\overline{w}, zw(h)), \\ l &= h - 1. \end{aligned}$$

The individual takes as given the earnings per effective labor schedule $w(h)$ and the function Γ_{ss} which determines the evolution of average lifetime earnings. In addition, the Social Security Administration does not tax earnings above $\widehat{y}$. Half of the social security taxes are paid by the employer and are not subject to the personal income tax τ_h.

3.6 Discussion on nonlinear earnings and labor supply decisions

The effects of nonlinear earnings on labor supply decisions can be illustrated with the following static labor-supply problem with no uncertainty and no heterogeneity ($z = 1$):

$$\max_{\{c,h\}} \left\{ \ln(c) + \varphi \frac{(1-h)^{1-\sigma}}{1-\sigma} \right\} \tag{17}$$
$$\text{s. t.}$$
$$c = w(h) + X,$$

where X denotes nonlabor income and $w(h) = \Theta h^{\frac{\varepsilon}{\theta}}$ for some constant Θ. The FOC necessary for an optimum solution is

$$MB(h) = \frac{1}{c} w'(h) = \frac{1}{\Theta h^{\frac{\varepsilon}{\theta}} + X} \Theta \frac{\varepsilon}{\theta} h^{\frac{\varepsilon}{\theta}-1} \le \varphi(1-h)^{-\sigma} = u_2(c, 1-h) = MC(h). \tag{18}$$

When earnings are a linear function of hours (the case of $\varepsilon = \theta$), the MB (marginal benefit) and MC (marginal cost) curves intersect only once and there is a unique local maximum (see Panel a in Figure 4). Since the second order conditions are satisfied, the local maximum is also a global maximum. When earnings are a nonlinear function of hours worked (the case of $\varepsilon > \theta$), the MC and MB curves intersect twice. Now there is a local minimum and a local maximum represented by points A and B on Panel b in Figure 4. The local maximum is not necessarily a global maximum as there maybe a corner solution with hours of work equal to zero $-$ this will occur when the area between the MC and MB curves to the left of point A is bigger than the area between the MC and MB curves between the points A and B, as illustrated on Panel b. A decrease in nonlabor income X shifts up the MB curve, making it less likely that a corner solution will arise since the area between the MC and MB curves to the left of A decreases while the area between the MC and MB curves between points A and B increases.

In the dynamic model with idiosyncratic shocks z, the FOC on labor hours at age j implies:

$$\ln(1-h_j) \ge -\frac{1}{\sigma} \left\{ \ln(1-\tau) + \ln(z_j w'(h_j)) + \ln \mu_j - \ln \varphi \right\} \text{ with equality if } h_j > 0, \tag{19}$$

where μ_j denotes the marginal utility of wealth at age j, z_j $-$ the age-j productivity shock, and τ $-$ the tax rate on earnings. Note that for a fixed μ_j, the FOC on labor supply (19) is the one in the static problem (see equation (18)). Hence, as in the static labor supply problem, the dynamic problem can feature a corner solution, a local minimum, and a local maximum. Assuming an interior solution in the labor supply decisions, using the Euler equation, and the fact that, under rational expectations, innovations to the marginal utility of wealth at age

$j+1$ (ϵ_{j+1}) should be uncorrelated with μ_j gives[15]

$$\Delta \ln(1-h_{j+1}) = -\frac{1}{\sigma}\left\{\Delta \ln(z_{j+1}w'(h_{j+1})) - \ln(\beta R_j) + \frac{\beta R_j \epsilon_{j+1}}{\mu_j} + \Delta\epsilon_{j+1}\right\}. \qquad (20)$$

Because the hourly wage $\left(w_h(h) = \frac{zw(h)}{h}\right)$ is proportional to marginal earnings $\left(zw'(h) = \frac{\varepsilon}{\theta}w_h(h)\right)$, it follows that

$$\Delta \ln\left(z_{j+1}w'(h)\right) = \Delta \ln\left(w_h(h)\right). \qquad (21)$$

Combining (20) and (21) gives a Frisch-elasticity of leisure $\eta^l = -\frac{1}{\sigma}$, which is equal to the expression obtained in a model with linear earnings. Hence, nonlinear earnings do not affect the Frisch-elasticity of labor supply along the intensive margin (see Aaronson and French (2009) for a related discussion.)

4 Calibration

The calibration of the model requires taking a stand on the value of the parameter σ determining the intertemporal elasticity of substitution of leisure. As we shall see, this parameter is important for how labor supply varies over the life cycle and how it responds to productivity shocks. Rather than calibrating σ, our approach is to fix exogenously a range of values for the parameter σ and to calibrate the model economy for each of these values. We then evaluate how the calibrated model economies match the facts.

The crucial task in our calibration is the parameterization of the stochastic process on labor productivity through an indirect inference approach. The estimation of this process exploits the fact that our theory of nonlinear earnings provides a very natural way to identify the measurement error in hours in the data. The calibration sets the model period to 1 quadrimester (a 4-month period) in order to model the variation in employment within a year. This choice allows us to use employment data, such as the fraction of individuals working all three quadrimesters in a year, from the Survey of Income and Participation Program (SIPP), which interviews individuals three times in a year (rather than once a year as in the PSID). The SIPP, however, is longitudinally fairly short to allow us to estimate the stochastic process for wages. As a result, we use the PSID for this purpose. In calibrating a *quadrimesterly* stochastic process on labor productivity, one difficulty arises from the fact that the PSID only reports earnings and hours of work at an annual frequency. Moreover, in using wage data to calibrate a stochastic process on labor productivity we need to take a stand on how hours of work affect labor productivity, and we need to consider that the data only report wages for

[15] $\mu_{j+1} = E_j\mu_{j+1} + \epsilon_{j+1}$

individuals that work. To deal with these problems, we follow an indirect inference approach (see Smith (1990), Gourieroux et al. (1993), and Guvenen and Smith (2010)):

1. Estimate an annual wage profile and wage process for college and non-college workers from the PSID data.
2. Use estimates from Aaronson and French (2004) on nonlinear earnings to pin down the parameter ε determining how hours of work affect labor productivity in the model economy.
3. Feed a *quadrimesterly* labor productivity process into the model economy.
4. Simulate the model economy to obtain *quadrimesterly* data on employment, hours of work, and earnings.
5. Aggregate the *quadrimesterly* data to an annual period.
6. Estimate an annual wage profile and wage process for college and non-college workers in the model generated data.
7. Feed a new *quadrimesterly* labor productivity process (go back to step 3), until the "same" annual wage profile and wage process is obtained in the model and in the data.

Below we describe the calibration in detail. We first discuss the calibration of the "macro" parameters and then proceed with a discussion of the calibration of the labor productivity process and how we deal with the possibility of measurement error in hours and earnings in the PSID data.

4.1 Calibration of preferences, technology, and macro parameters

The model period is set to one quadrimester (4 months). The model economy is solved in partial equilibrium for a fixed interest rate. The quadrimesterly interest rate is chosen so that the implied annual rate of return on capital (net of depreciation) is 4%.[16]

The intertemporal elasticity of substitution of leisure. We calibrate model economies with $\sigma = 1.8$, 2.0, and 2.2 so that the intertemporal elasticity of leisure varies from 0.56 to 0.45 across the calibrated model economies. We also attempted to calibrate model economies

[16]The depreciation over a yearly period is assumed to be 4%. Because the model economy is solved in partial equilibrium, the depreciation rate does not affect any of the results in the next section of the paper.

with values of σ below 1.8 and above 2.2 but these economies turn out to match quite poorly the calibration targets and the rest of the micro level facts.

Preference parameters, time endowments, and mortality rates. Following Kaplan and Violante (2010), the discount factor β is chosen to match an asset to income ratio of 2.5. This is the wealth to income ratio when the top 5% of households in the wealth distribution are excluded from the Survey of Consumer Finances. The reason for excluding the richest households in computing an aggregate wealth to income ratio is that the PSID undersamples the top of the wealth distribution. Following Osuna and Ríos-Rull (2003) and Prescott (2004), the time endowment is set at 5200 hours a year (100 hours per week). The preference parameter φ, determining the taste for leisure, is chosen so that prime age individuals, aged 30 to 45, work annually about 42% of their available time. The mortality risk for college and non-college individuals is taken from Bhattacharya and Lakdawalla (2006).

Technology parameters. The labor share θ is set to 0.64. In order to calibrate the parameter ε, we use the fact that the equilibrium hourly wage in our theory satisfies

$$w_h(h) = \frac{zw(h)}{h} = \Theta h^{\frac{\varepsilon}{\theta}-1}. \tag{22}$$

Note that the elasticity of the wage rate to a change in hours of work is given by $\frac{\varepsilon}{\theta} - 1$. In an empirical study, Aaronson and French (2004) estimate this elasticity to be around 0.40. This estimate implies that a full time (40 hours a week) worker earns an hourly wage 25% higher than a part time (20 hours) worker. We thus set $\varepsilon = 1.4\ \theta$.

Tax rates and social security. The tax rate on consumption τ_c is set at 0.055 as in Conesa et al. (2009). Following Domeij and Heathcote (2004), taxes on capital income and labor income are set to $\tau_k = 0.40$ and $\tau_h = 0.27$. The social security tax rate is set to $\tau_{ss} = 0.12$, and the cap $\widehat{y}$ on social security taxation is fixed at 2.47 of average earnings in the economy ($\overline{W}$).[17]

4.2 Calibration of labor productivity

We use a GMM procedure to estimate the following annual wage process in the PSID data for college and non-college individuals:

$$\ln(\widehat{w_h})_{ij} = x_j\kappa + \alpha_i + u_j + \lambda_j, \tag{23}$$

[17] Actually, $\overline{W}$ is set at 80% of average earnings in the economy. The reason is that our model only includes male workers. Using data from the CPS, we find that the average earnings among all workers in the US economy are about 80% of the average earnings of male workers.

where $\ln(\widehat{w_h})_{ij}$ represents the observed log hourly wage of individual i at age j in the PSID data, x_j is a quartic polynomial in age, κ is a vector of coefficients, $\alpha_i \sim N(-\frac{\sigma_\alpha^2}{2}, \sigma_\alpha^2)$ is a fixed effect determined at birth, $\lambda_j \sim N(-\frac{\sigma_\lambda^2}{2}, \sigma_{\lambda j}^2)$ is an idiosyncratic transitory shock which is age-dependent, and u_j follows a first-order autoregression:

$$u_j = \rho u_{j-1} + \eta_j, \quad \eta_j \sim N(-\frac{\sigma_\eta^2}{2}, \sigma_\eta^2), \quad u_0 = 0. \tag{24}$$

While the parameters $(\kappa, \rho, \sigma_\alpha^2, \sigma_\eta^2, \sigma_{\lambda j}^2)$ vary across education types, this is omitted in the notation to avoid clutter. The estimated wage processes are reported in Table 2. The empirical findings show that the variance of the fixed effects is quite large for both education types, with values of 0.097 and 0.072 for the non-college and the college types. Both wage processes exhibit high autocorrelation, with a value of 0.94 for non-college individuals and 0.98 for college individuals. The variance of the innovation of the autoregressive process is 0.019 and 0.021, respectively. The estimates reveal that both education types exhibit transitory shocks to wages with quite high variances as reported in Figure 7.

In order to calibrate the model economy, we need to find a *quadrimesterly* stochastic process on labor productivity that is consistent with the annual wage process estimated in the data (equations (23)-(24)). To do this, we assume that labor productivity is the sum of an annual autoregressive process and a quadrimesterly transitory shock.[18] Specifically, while the transitory shock is drawn every quadrimester, the persistent shock is only drawn at the first quadrimester of each year (age). To make these assumptions operational, we discretize all shocks by considering, for each education type, 15 values for the autoregressive shocks, 4 values for the temporary shocks, and 2 values for the fixed effects. The transition probabilities of the persistent shock are computed using a Tauchen (1986) routine.

The empirical literature has stressed the importance of measurement error in hours and earnings in household survey data. Moreover our empirical findings are suggestive of the importance of measurement error since the estimated variation in the transitory component of wages seems implausibly large (see Figure 7).

We thus need to address the issue of measurement error in the data. To this end, we assume that the transitory shock λ_j in the empirical model is the sum of a true temporary log wage shock and measurement errors in log hours m_H and log earnings m_E, both of which are normally distributed with mean zero. The estimated transitory variation in *observed*

[18] We have also experimented with a specification that allows for an autoregressive process at the quadrimesterly level. In this case, however, we were not able to recover the stochastic process estimated in the data. When labor productivity follows an autoregressive process at the quadrimesterly level, there is no reason to expect the logarithm of the sum of quadrimesterly earnings to be well approximated with an autoregressive process.

transitory log wages $\sigma_{\lambda j}$ is then the sum of the variances of transitory true log wages σ^2_{Tj}, measurement error in log earnings σ^2_{Ej}, and measurement error in log hours σ^2_{Hj}:

$$\sigma^2_{\lambda j} = \sigma^2_{Tj} + \sigma^2_{Ej} + \sigma^2_{Hj}. \tag{25}$$

We use the implications of our theory in order to take a stand on the relative importance of $(\sigma^2_{Tj}, \sigma^2_{Ej}, \sigma^2_{Hj})$ in accounting for the estimated variance in observed transitory log wages $\sigma^2_{\lambda j}$. Thus, we assume that annual hours and earnings are measured with error in the model economy. To calibrate the variance of true transitory log wages σ^2_{Tj}, we note that in our theory this variance has important effects on the probability that individuals work all three quadrimesters in a year: the larger σ^2_{Tj} is the less likely it is that individuals will work during all periods in a year. We thus use this statistic as a calibration target where the fraction of individuals working 3 quadrimesters in a year is taken from the Survey of Income and Participation Program (SIPP).[19] Figure A-4 shows that this fraction is roughly constant for prime-age males (age 30 to 50) but that it decreases substantially after age 50. To mimic the data in a simple way, for each education group the process for transitory shocks is parameterized with two values σ^2_{T50} and σ^2_{T64}, where the variance of the transitory shocks is assumed to be equal to σ^2_{T50} up to age 50 and then increasing linearly up to the value σ^2_{T64}.[20]

To distinguish between σ^2_{Ej} and σ^2_{Hj}, we need an additional target.[21] This is done by comparing the variance of transitory wages in two alternative estimations of the wage process in (23)-(24). The first specification estimates the process for observed wages while the second specification estimates the process for wages net of the effect of hours worked on wages. Identification comes from the fact that measurement errors in hours and in earnings affect differently the variance of transitory wages in the two specifications of the regression. To develop this point, we start by noticing that when earnings are a nonlinear function of hours, the observed hourly wage is given by

$$\widehat{w_h} \equiv \frac{\widehat{zw(h)}}{\widehat{h}} = \frac{\Theta z h^{\frac{\varepsilon}{\theta}} e^{m_E}}{h e^{m_H}} = \Theta z h^{\frac{\varepsilon}{\theta}-1} e^{m_E - m_H}, \tag{26}$$

where $\widehat{x}$ denotes the observed value for the variable x in the data. In the absence of measurement error, the wage rate net of the effect of hours on wages would be uncovered by taking

[19] We note that the SIPP allows us to have more reliable measures of labor force participation at the quadrimesterly frequency than the PSID as it interviews individuals three times in a year (rather than once a year as in the PSID).

[20] The value of σ^2_{T50} is 0.0179 for non-college and 0.0148 for college while σ^2_{T64} takes the value 0.0151 for non-college and 0.0097 for college.

[21] In the following discussion, in order to keep the notation compact, we abstract from the variation in σ^2_{Ej} and σ^2_{Hj} by age.

logs and subtracting $(\frac{\varepsilon}{\theta}-1)\ln h$ from both sides of (26):

$$\ln \widehat{w_h} - \left(\frac{\varepsilon}{\theta}-1\right)\ln h = \ln \Theta z.$$

In practice, though, hours are observed with error. Subtracting $\left(\frac{\varepsilon}{\theta}-1\right)\ln(he^{m_H})$ from both sides of equation (26) to "clean" wages from the effect of observed hours gives

$$\ln \widehat{w_h} - \left(\frac{\varepsilon}{\theta}-1\right)\ln(he^{m_H}) = \ln \Theta z + \left(\frac{\varepsilon}{\theta}-1\right)\ln h + m_E - m_H - \left(\frac{\varepsilon}{\theta}-1\right)\ln(he^{m_H}),$$

which can be re-arranged as

$$\ln \widehat{w_h} - \left(\frac{\varepsilon}{\theta}-1\right)\ln(he^{m_H}) = \ln \Theta z + m_E - \frac{\varepsilon}{\theta}m_H. \tag{27}$$

If z follows the empirical model in (23)-(24), then we obtain the following empirical model for "clean" wages:

$$\ln \widehat{w_h} - \left(\frac{\varepsilon}{\theta}-1\right)\ln(he^{m_H}) = \ln \Theta + x_j\ \kappa + \alpha_i + u_j + \lambda_j + m_E - \frac{\varepsilon}{\theta}m_H \tag{28}$$

The transitory variation in "clean wages" is then given by

$$VAR(\lambda_j + m_E - \frac{\varepsilon}{\theta}m_H) = \sigma_T^2 + \sigma_E^2 + \left(\frac{\varepsilon}{\theta}\right)^2\sigma_H^2. \tag{29}$$

When earnings are a nonlinear function of hours worked ($\frac{\varepsilon}{\theta} > 1$) and the wage process is estimated net of the effect of hours on wages, measurement error in hours leads to an increase in the estimated transitory variation of wages. Intuitively, this happens because we are not using the "correct" hours to clean the wage data. Comparing (29) with (25), the increase in the variance of transitory wages is given by

$$\Delta VAR = \left[\left(\frac{\varepsilon}{\theta}\right)^2 - 1\right]\sigma_H^2. \tag{30}$$

For the calibrated value of $\frac{\varepsilon}{\theta} = 1.4$, we have that $\left[\left(\frac{\varepsilon}{\theta}\right)^2 - 1\right] \simeq 1$ so that $\Delta VAR = \sigma_H^2$. Thus, for each education type, the variance of the measurement error in hours σ_H^2 is obtained as the increase in the transitory variance in wages when the wage data is "cleaned" with hours data. We then introduce the estimates for measurement error in hours into the model economy and run the two specifications of the wage regression with model simulated data. Reassuringly, as discussed above, when the estimation is performed on clean wages the variance of the transitory component in wages increases by an amount approximately equal to the measurement error in hours estimated in the data.

4.3 Calibration results

Overall, we find that there is a "tight" range of values of σ (going from 1.8 to 2.2) for which the calibrated model economies match well the calibration targets and other relevant micro facts. Hence, below we set the economy with σ equal to 2 as our baseline economy and discuss in detail its performance. While the analysis mostly focuses on the baseline economy, we compare how the parameter σ affects the individual and aggregate labor supply responses across the calibrated model economies.

Unless otherwise indicated, we only report the results for the baseline economy ($\sigma = 2$). We calibrate the model parameters by solving the model economy. Table 3 shows the values and the calibration targets for three of these parameters: the average earnings in the economy $\overline{W}$, the taste for leisure φ, and the discount factor β. For each education group, we use an indirect inference approach to pin down a quartic polynomial for the wage age-profile, the stochastic process of wages, and the variance of measurement error in hours and earnings.

4.3.1 Wages: age profile and stochastic process

Since in our baseline economy there is an active extensive margin in labor supply decisions, individuals who work are a non-random selection of the population. Hence, we cannot mechanically plug an age-profile for wages into our model. Nonetheless, Figure 5 shows that the baseline economy matches almost exactly the age-profile of wages for both education groups. Table 4 reports the values of the parameters characterizing the AR(1) process as well as the standard deviation of the fixed effect shock affecting labor productivity for non-college and college types. This table also reports the targeted statistics which are the estimated variance of the fixed effect and the parameters of the AR(1) process for log wages (also in Table 2). The values reported under the column Model correspond to the GMM estimation using annual model data from the baseline economy.

As explained in the calibration procedure, the process for transitory shocks is parameterized with two values σ^2_{T50} and σ^2_{T64}, where the variance of transitory shocks is assumed to be equal to σ^2_{T50} up to age 50 and that it then changes linearly to the value σ^2_{T64}. The targets are the fractions of prime-aged males (age 35-50) and males aged 50-64 that work all three quadrimesters in a year. The calibration results in σ^2_{T50} equal to 0.0179 for non-college and 0.0148 for college individuals, while σ^2_{T64} takes the value 0.0151 for non-college and 0.0097 for college. Hence, to match the fact that in the data non-college individuals are less likely to work three quadrimesters in a year, the calibration implies that non-college individuals are subject to larger true temporary shocks than college individuals. The model replicates reasonably well the fraction of people working 3 quadrimesters in a year, though for young non-college individuals the calibration tends to slightly overpredict the fraction of individuals working all periods in a year (see Figure 6).

4.3.2 Measurement error

To estimate the measurement error in hours our calibration procedure compares the variance of transitory wages in two alternative estimations of the wage process both in actual and in model data. The first specification estimates the wage process using data on observed wages while the second specification estimates the process for wages net of the effects of hours worked on wages. Regarding the first specification, Figure 7 shows that the model matches well the age-profile for the variance of the transitory component of residual log wages in the data for both education groups. Figure A-5 shows that the model also matches well the variance of the transitory component for "clean" wages. Measurement error in hours is obtained as the difference between the variance of transitory wages across the two specifications for the wage process. The results for measurement error in hours are reported in Figure A-6. We find that measurement error in hours is higher for non-college than for college individuals. The age-profile of measurement error is slightly U-shaped for non-college individuals: the variance takes a value of around 0.03 for very young individuals, is below 0.03 for prime-aged males, and increases to 0.05 when individuals are close to the retirement age. For college individuals, the variance in measurement error is about 0.02 for most of the life cycle, with a mild increase to 0.04 prior to retirement.

5 Quantitative findings: the performance of the baseline economy

5.1 Hours worked

Figures 8-12 present the performance of the baseline economy in accounting for the facts on labor supply documented in Section 2. Overall, the baseline economy captures most of the salient features of labor supply. Recall that the facts on labor supply were not explicitly targeted, indicating that the features included in the analysis are important determinants of individuals' labor supply decisions.

Age profile of hours of work. Figure 8 displays mean annual hours over the life cycle in the model and for various cohorts in the data. The model captures very well the fact that working hours over the life cycle are roughly flat up to age 45 and that they decrease steeply after age 45 (though it does not account for all the decline in hours late in the life cycle). The model matches fairly well the decline in working hours among individuals with positive hours of work as shown in Figure 9. Our baseline economy with an intertemporal elasticity of leisure of 0.5 accounts for the low comovement of hours and wages early in the life-cycle, when wages are rising rapidly but hours are relatively flat. Incomplete markets are crucial for this result.[22] While individuals face an increasing age profile of wages, they work long hours when young because they need to build a buffer stock of savings to self-insure against income risk. By age 50 the stock of assets is sufficiently large, and individuals can afford to take a quadrimester off work when they receive a low realization of the temporary wage shock. This accounts for the pronounced decline in annual working hours late in the life cycle.

Table 5 presents and summarizes the cross-sectional distribution of labor supply in the model and in the data. We capture the cross-sectional distribution of labor supply by restricting the analysis to a quadrimesterly level both in the model and in the SIPP data. In order to analyze the extensive margin, we use a probit regression of employment on age, education, and assets. For the analysis on the intensive margin we use an OLS regression of log hours on age, education, and assets.[23] Age ranges from 25 to 65, assets are in thousands, and education is a dummy variable which takes the value of one if the individual is a college graduate and zero otherwise.[24]

The results in Table 5 indicate that labor supply both on the intensive and the extensive margin declines with age.[25] Quantitatively, this decline is quite similar in the model and in the data. The model also quantitatively captures the fact that college graduates work more − both on the intensive and the extensive margin − even though this effect is less pronounced in the model than in the data. Finally, while it seems that assets have a positive effect on labor supply at the intensive margin and a negative effect at the extensive margin, these effects are quantitatively very small both in the model and in the data. For instance, an increase in

[22] Imai and Keane (2004) develop a theory with human capital accumulation that accounts for the comovement of hours and wages early in the life cycle with an intertemporal elasticity of labor supply of 3.8. However, this high intertemporal elasticity results in hours worked declining much too steeply relative to the data in the second half of the life cycle, a point recently shown by Wallenius (2009).

[23] We are using the 1990 SIPP Panel in the analysis. Assets data is reported only for wave 4 (out of 8), and we use the information only at that point. We construct two series for assets: the first includes both own and joint assets while the second includes own assets plus half of the joint assets. SIPP writes the joint asset information either in the husband's or the wife's record, but not both. We match couples in order to obtain the information about joint assets. Assets are classifies into 6 large groups: house, mobile home, other real estate, vehicles, financial assets, and rental properties and other real estate. We use net assets by subtracting liabilities. Further, the reported results are obtained using only the financial assets information (liquid net wealth), excluding mortgages and Individual Retirement Accounts.

[24] The analysis on the SIPP data also includes a race dummy variable.

[25] Table A-6 reports the results when the analysis is performed separately for college and non-college.

assets by \$10,000 decreases the probability of being employed by 0.3% in the model and 0.06% in the data. It seems that in this simple correlation analysis, once we take into account the effect of age and education, the additional effect of assets on labor supply is small. It is indeed the case that individuals with more assets have a lower probability of being employed, but in the regression analysis this is mainly captured by the age variable since older individuals tend to be also wealthier.

We emphasize that life cycle is important for generating a low correlation between assets and non-employment spells. When the time horizon is infinite and labor productivity follows an autoregressive process, an individual receiving a low labor productivity shock understands that his labor productivity will eventually revert to the mean. His optimal decision is then not to work as long as consumption can be financed with non-labor income. Since forward looking individuals understand that they are likely to receive low productivity shocks in the future, they build precautionary savings in order to finance the non-employment spells during periods of low labor productivity. This mechanism explains why Castaneda et al. (1998) find that non-employed individuals counterfactually tend to be too asset rich in their infinitely lived framework (see also Chang and Kim (2006)). On the other hand, in a life-cycle model the finite horizon makes individuals less willing to take long spells out of employment for two reasons. First, young individuals do not have a buffer stock of assets to finance non-employment (at least for long spells). Second, a finite life implies that middle-aged and old individuals, who are asset rich, may start receiving high productivity shocks too late in their life cycle and thus may not be able to benefit from the good realizations of the productivity shocks or benefit from them for only a short period of time.

Dispersion in hours of work. Figure 10 shows the dispersion of annual hours worked over the life cycle both in the model and in the data. For both education types, the model captures the fact that the dispersion in working hours is flat for prime-age males and that it increases substantially late in the life cycle. Again, incomplete markets account for these findings. When individuals are young, the dispersion in hours is low because most individuals are working long hours to self-insure against income risk. After age 45, most individuals have a buffer stock of savings that allows them to take some periods off work and the dispersion in working hours rises substantially. The baseline economy underpredicts the dispersion in hours worked but this should not be surprising as the model abstracts from many factors that could lead to heterogeneity in working hours across individuals. In considering what these factors may be, it is suggestive that the baseline economy underpredicts the heterogeneity in lifetime labor supply (see Table 1). While in the data the coefficient of variation in lifetime labor supply (labor supply over a ten year period) for individuals aged 35 to 45 and aged 45 to 55 is 0.26 and 0.37, this statistic takes values of 0.11 and 0.16 in the baseline economy for the two age groups considered. Hence, it seems that the theory abstracts from some factors leading to persistent differences in working hours across individuals in the U.S., such as heterogeneity in health, preferences, or demographics. This observation is supported by the findings of Bils et al. (2009) who model and calibrate permanent differences in tastes for work across individuals to match micro facts in the US economy.

Persistence in annual hours worked. Following the data analysis in Section 2, we divide individuals in the baseline economy into four groups according to their annual working hours: The first group corresponds to individuals who do not work (annual hours less than 100); the second group is given by individuals who work part time (annual hours between 100 and 1500); the third group corresponds to people working full time (annual hours between 1500 and 2800); and the fourth includes individuals working overtime (annual hours greater than 2800). Figure 11 shows the relative size of each of these groups over the life cycle both in the model and in the data. Two observations stand out. First, the model captures the fact that Group 3 (people working between 1500 and 2800 hours) is by far the largest group and that its share only declines significantly after age 55. Second, the model mimics the observation that the size of Group 1 (individuals working between 0 and 100 hours) in the data is very small for people younger than 50, but then rises substantially late in the life cycle.

Figure 12 shows the persistence in annual hours worked both in the model and in the data. First, the model captures the fact that individuals working between 1500 and 2800 hours in a year (Group 3) are quite likely to be in the same group the year after. Second, the model also mimics the observation that early in the life cycle people who do not work in a given year tend not to stay in the same group the year after. Later in the life cycle, however, non-participation becomes an absorbent group since non-working individuals tend to stay in that group with a very high probability. Finally, the other two groups are not very big and tend to be transitory − individuals end up in those groups every now and then, but tend to quickly exit them.

Distribution of hours. Figure 13 compares the distribution of hours in a quadrimester both in the model and in the data for two education groups and three age groups. The model mimics the facts that the distribution of hours is highly concentrated around 600 hours and has a spike at zero. As previously discussed, the fraction of individuals working zero hours increases with age because older individuals are richer and can afford to take periods off work. The fraction of people working zero hours is highest for non-college individuals than for college individuals, especially for the older age groups. This is explained by the fact that non-college individuals have a flatter age profile of wages and face (slightly) lower transitory shocks to wages. For all age and education groups the model underpredicts the fraction of individuals working zero hours.

5.2 Labor supply responses at the individual level

We have shown that the baseline economy matches the calibration targets plus other relevant micro facts on labor supply. Now, we evaluate the predictions of the theory for changes in hours and wages at the individual micro level.

5.2.1 Changes in log hours and wages: covariance and correlation

Table 6 shows that both in the baseline economy and in the data the covariance between changes in log hours and changes in log wages have an inverted U-shape and a negative sign. The baseline economy matches these data remarkably well. Table 7 shows that the baseline economy is also successful in predicting the negative correlation between the change in log hours and the change in log wages in the data, though the quantitative fit is not as good as in the case with the covariances between changes in log wages and log hours.

It is interesting to compare labor supply responses across the calibrated model economies. While the covariance between changes in hours and wages does not vary much across economies, it tends to increase with the value of σ. This result is explained by the fact that our calibration implies a monotonic relation between σ and the variance of the transitory shocks. Recall that the calibration pins down the variance of transitory shocks in order to match the fraction of individuals working three quadrimesters in a year. Note that an increase in σ, ceteris paribus, makes individuals less willing to work long hours because of the higher curvature of the utility function. Hence, to match the target for working hours (42%) the calibration of the model requires that the weight of leisure (φ) in the utility function decreases with σ. A decrease in the value of leisure φ also implies that individuals are less willing to take periods off work, thereby implying that the variance of transitory shocks should increase with σ in order to match the calibration target for the fraction of people working 3 quadrimesters in a year. Since transitory shocks have a small wealth effect, the increase in the variance of transitory shocks leads to a higher covariance of hours and wages.

5.2.2 The empirical elasticity of labor supply

While the majority of empirical studies on the labor supply of men estimate low elasticities, there is no clear consensus on the magnitude of this elasticity. Furthermore, it is well known that there are many serious econometric problems in estimating the elasticity of labor supply (see Keane (2010) and Keane and Rogerson (2011) for a recent survey of the literature). Nevertheless, our model economy is consistent with a very low empirical elasticity of labor supply. We simulate micro data from the model economy and estimate an "empirical elasticity of labor supply" with standard econometric techniques. We find that the empirical estimates in the model simulated data imply a very low empirical Frisch elasticity of labor supply − depending on various specifications it is in the range of $[-0.07, 0.36]$, which is well within the range of $[0, 0.5]$ reported in various empirical studies.[26] Similar to the findings in Imai and Keane (2004), Domeij and Flodén (2006), and Rogerson and Wallenius (2009) the estimated elasticities are lower than the Frisch elasticity of labor supply hard-wired in the calibration of the model economy (i.e., the "theoretical elasticity of labor supply"). As discussed in Section 3.6, the Frisch elasticity of leisure along the intensive margin in the model with nonlinear earnings is $\eta^l = -\frac{1}{\sigma}$, which is the same as in a model with linear earnings. It is standard to convert the elasticity of leisure into a labor supply elasticity by setting $\eta^h = -\frac{(1-h)}{h}\eta^l$. Hence, the theoretical Frisch elasticities of leisure and labor in the baseline economy are $\eta^l = -0.5$ and $\eta^h = 0.61$.[27]

6 Aggregate labor supply responses

The analysis so far has demonstrated that the model developed earlier is able to match quantitatively quite well the facts on labor supply at the individual level. Therefore, it is an appropriate tool for studying aggregate labor supply responses to changes in the economic environment. The model allows us to explicitly aggregate up from each individual's response, as well as analyze the response in various parts of the age, education, asset, and productivity shocks distribution.

We start by studying the aggregate labor supply response to two different types of changes in the economic environment. The first experiment involves a one period unanticipated wage change. The wealth effect of such a change is negligible, and we can use the corresponding

[26] In the Appendix we describe the procedure in greater detail, and we also discuss the importance of time aggregation and the extensive margin on the estimated low empirical labor supply elasticity.

[27] To map the elasticity of leisure into an elasticity of labor, we use that the median value of $(1-h)/h$ in our model economy is 1.22.

change in aggregate labor supply in order to obtain an estimate of its Frisch elasticity. The second experiment involves a permanent (compensated) wage change in which we permanently increase the labor income tax rate and redistribute back the tax receipts. This experiment is designed to provide a measure of the compensated Hicksian elasticity of aggregate labor supply. Furthermore, we also simulate the 1987 tax holiday in Iceland − a quasi natural experiment − and find that the aggregate labor supply responses in the model are similar to those actually observed in Iceland and reported in Bianchi et al. (2001). Finally, we discuss the fact that the model is also consistent with the cross-country evidence on labor supply late in the life cycle (after the age of 50) and analyze the effect of nonlinear earnings on the results.

6.1 The Frisch elasticity of aggregate labor supply

We simulate a one-period (quadrimester) unanticipated wage increase of 2 percent. The results are reported in Table 8. We find that the Frisch elasticity of aggregate labor supply in our baseline economy is 1.27, which is twice as big as the theoretical Frisch elasticity embedded in the calibration of the model (0.61). The large labor supply response along the extensive margin explains why the Frisch elasticity is much bigger than the theoretical elasticity (recall that the latter was derived assuming an interior solution in the labor supply decision.) Restricting attention to labor supply changes along the intensive margin decreases the Frisch elasticity from 1.27 to 0.58. Hence, the extensive margin accounts for about 54% of the aggregate labor supply response to a temporary wage change.

Not surprisingly, the Frisch elasticity of aggregate labor supply in the model economy is sensitive to the value of σ determining the i.e.s.: as σ varies in the tight range between 1.8 to 2.2, the Frisch elasticity decreases from 1.56 to 1.07 (see Table 8). Table 9 reports the elasticity of labor supply for different age groups in the baseline economy. The elasticity of labor supply increases steeply with age: it rises from 1.0 for individuals aged 25-35, to 1.98 for individuals aged 55-64. While the response along the intensive margin is roughly flat over the life cycle, the wage elasticity of employment rises from 0.38 for individuals aged 26-35 to 1.56 for people aged 55-64. The employment of old individuals is very responsive to temporary wage changes because they, on average, have a buffer stock of savings that allows them to smooth consumption well in response to an unanticipated wage shock. On the contrary, young individuals are less responsive in their labor supply because they are poorer and they need to build a stock of precautionary savings to insure against income risk over the life cycle.

Unfortunately, there is no direct micro evidence on the wage elasticity of employment to unanticipated temporary shocks (or over the business cycles). The key difficulty is that wages are only observed for employed people. To deal with this missing wage problem, Kimmel and Kniesner (1998) extend the Heckman and MaCurdy (1980) analysis to estimate a labor supply equation jointly with a participation decision rule and an offer wage function at sub-annual periods. They use SIPP data and instrument for wage rates using nonlinear age and time trends. While these instruments are not beyond dispute, they find that the extensive margin accounts for almost 70% of the wage elasticity of labor supply, with an estimate of the wage elasticity of labor supply of 1.25 and a wage elasticity of employment of 0.86.[28] These estimates are quite close to our findings for a one period small wage change (aggregate elasticity of 1.27, with the extensive margin contributing 0.69). Our theory calibrated to micro data on hours and wages is consistent with a Frisch elasticity of aggregate labor supply well above 1. Nonetheless, our estimate for the employment elasticity is well below the value of 2 needed to match the business cycle data along the extensive margin, as argued by Chetty et al. (2011). These authors claim that it is important to model unemployment or demand-driven movements on the employment rate for understanding employment fluctuations over the business cycle. Our results are consistent with this view.

6.2 The Hicks elasticity of aggregate labor supply

To evaluate the compensated elasticity to a permanent wage change, we simulate an increase in the labor income tax from 0.27 in the baseline economy to 0.37. Tax proceeds are assumed to be rebated with a lump sum transfer to working-age individuals. The amount of the transfer is education-specific so that there is no income redistribution across education types. We find that the elasticities for both the intensive and extensive margins are reduced by half relative to the case of a temporary wage change (see Table 8). Now the labor supply elasticity is 0.65 and the employment elasticity is 0.35, which should be compared to 1.27 and 0.69 − the elasticities from the temporary wage change experiment. Hence, not surprisingly, individuals respond more strongly to a temporary wage change than to a permanent compensated-wage change. The age profile of the wage elasticity to the permanent change in wages has a U-shape. It starts at 0.70 for young individuals, decreases to 0.56 for individuals aged 55-64 and it increases to 0.72 for people aged 55-64 (see Table 9). Young individuals respond strongly to the tax increase because the lump sum transfer helps them to smooth consumption. This effect is stronger for college than non-college individuals since college individuals face a steep age profile of labor productivity. As a result, the aggregate elasticity for the age group 25-35

[28]The exclusion restriction is that employment rates do not vary with age conditional on wage rates. Chetty et al. (2011) argue that the elasticity estimates would be biased upwards if factors that predict high wage rates also predict high latent tastes for work.

is higher for college than for non-college individuals (0.81 versus 0.65).

The mean value of the Hicks elasticity of aggregate hours across a number of micro studies reviewed in Chetty et al. (2011) is 0.76. Consistent with this evidence, our baseline economy predicts a Hicks elasticity of aggregate hours of work of 0.65. Since this finding also holds in Rogerson and Wallenius (2009), we conclude that modeling heterogeneity and incomplete markets is not crucial for the predictions of the theory for the Hicks elasticity of labor supply.

6.3 The Icelandic tax holiday experiment

We simulate the tax holiday that took place in Iceland in 1987. The Icelandic tax holiday is well suited for identifying Frisch elasticities because it induced an unanticipated temporary wage variation during the year 1987. In 1987, Iceland moved from a system under which taxes were paid on the previous year's income to a pay-as-you-earn system. The transition to the new tax system implied that income during 1987 was never taxed since the tax base in 1987 was income earned in 1986 and the tax base in 1988 was income earned in 1988.[29] The average tax rate was 14.5% in 1986, 0% in 1987, and 8.0% in 1988. In order to mimic the tax reform in Iceland, we simulate in our baseline economy a one year (three model periods) reduction in the tax rate of 14.5 percentage points, followed then by a permanent decrease of 6.5 percentage points in the average tax rate (relative to the initial tax rate of 27% in the baseline economy).

We find that the aggregate elasticity of labor supply implied by the Icelandic tax holiday experiment is 0.68, with extensive and intensive margin elasticities of 0.33 and 0.35, respectively. Remarkably, these elasticity results are not much different from the ones estimated by Bianchi et al. (2001) in the Icelandic micro data: for male workers, they find an employment elasticity of 0.58 and an intensive margin elasticity of 0.26.[30]

We find much smaller elasticities than the ones reported in Chetty et al. (2011) obtained by simulating the Icelandic tax holiday in the Rogerson and Wallenius (2009) model. This is due to the fact that our baseline economy with heterogeneous agents and incomplete markets has a smaller fraction of agents that are close to being indifferent between working or not working than in the Rogerson and Wallenius (2009) representative agent model.

The Icelandic tax holiday experiment assumes a temporary wage change that lasts three model periods (one year). As we discussed earlier, simulating the effects of a one period

[29]The tax change was unanticipated by households since the announcement of the policy change was made in late 1986, see Bianchi et al. (2001).

[30]The extensive margin elasticity is reported in Table 4 in Bianchi et al. (2001). The intensive margin elasticity is obtained using data from Table 6 in Bianchi et al. (2001).

(quadrimester) wage increase of 2% delivers an aggregate elasticity of labor supply equal to 1.27, with an extensive margin elasticity of 0.69 (see Table 8). The extensive margin response is larger than the one obtained in the Icelandic experiment (0.33) for three reasons. First, the change in wages lasts only one period (quadrimester) rather than three model periods (one year) and the scope for intertemporal substitution is higher in the case of a one period wage change. Moving from a one quadrimester to a one year change decreases the employment elasticity from 0.69 to 0.61. Second, the employment response in our model is nonlinear in the size of the wage change: the employment elasticity to a one-year wage change drops from 0.61 to 0.41 when the size of the wage change increases from 2% to 18%. Third, the Icelandic tax experiment combines a one-year temporary change with a permanent wage change, which further decreases the employment elasticity from 0.41 to 0.33.

6.4 Intertemporal substitution and labor supply late in the life cycle

There are substantial cross-country differences in labor supply late in the life cycle (age 50+). In many countries the social security provisions impose explicit and implicit taxes on the labor earnings of individuals that are close to the normal retirement age, encouraging these individuals to reduce their labor supply and withdraw from the labor market before the normal retirement age (Gruber and Wise, eds (1999)). The cross-country heterogeneity in labor supply and social security provisions present us with the opportunity to test the predictions of our theory. In Erosa et al. (2012) we extend our baseline economy to model in detail the variation in the social security, disability insurance, and taxation institutions across European countries and the United States. We find that the extended baseline model economy accounts well for the observed cross-country differences in labor supply late in the life cycle, indicating that the Frisch elasticity of labor supply in our model economy is plausible. That the model economy would pass this test was not obvious: Imai and Keane (2004) estimate an intertemporal elasticity of substitution in labor supply of 3.82 in a structural model of human capital accumulation and labor supply decisions fitted to NLSY data on hours and wages of young individuals (aged 20-36). However, Wallenius (2009) shows that the estimated model of Imai and Keane (2004) counterfactually implies a too sharp reduction in working hours late in the life cycle (age 50+).

6.5 Discussion on nonlinear earnings

Nonlinear earnings and aggregate labor supply responses. A distinguishing feature of our theory is that nonlinear earnings imply that the aggregate elasticity of labor supply is substantially larger than the theoretical elasticity implied by the calibration of preference parameters. This is because the theoretical elasticity only describes labor supply responses along the intensive margin, thereby neglecting labor supply responses along the extensive margin. Imai and Keane (2004) and Domeij and Flodén (2006) are two closely related papers that argue that standard econometric estimates of the elasticity of labor supply may be biased downwards. However, these papers cannot account for large aggregate labor supply responses because they do not model the extensive margin.

Imai and Keane (2004) estimate a very large value for the intertemporal elasticity of substitution (i.e.s.) of labor − roughly about 4. The key to their large estimate is that the incentives to supply labor in their framework are driven by the sum of the wage rate and the returns to human capital accumulation. When the returns to human capital are large, labor supply responds little to wage changes leading to high estimates of the i.e.s. Notice, however, that this reasoning also implies that labor supply should not be very responsive to aggregate wage shocks. Indeed, Imai and Keane (2004) simulate the effects of a one-period change in the wage rate of 2% and find that the average labor supply change for individuals aged 30 to 50 is about 1.5%, a much smaller response than the 8% predicted by the estimated i.e.s. of 4. On the contrary, in our paper the labor supply response to a temporary wage change is more than twice the value predicted by the i.e.s. parameter. In our framework with nonlinear earnings, temporary wage changes have a large effect on the extensive margin leading to labor supply responses larger than those implied by the i.e.s. parameter. Our findings provide an explanation for the evidence in Kimmel and Kniesner (1998) of a highly wage-elastic extensive margin at subannual periods in the SIPP data.

Domeij and Flodén (2006) argue that standard econometric estimates of the elasticity of labor supply may be biased downwards due to the presence of liquidity constraints. The idea is that individuals who are liquidity constrained may not be able to reduce their labor supply when they are hit by a negative temporary wage shock. Since consumption smoothing can only be achieved by an increase in labor supply, the labor-supply response of liquidity constrained individuals is thus smaller or of the opposite sign that what is predicted by an analysis that ignores such constraints. Note, again, that this reasoning also implies that labor supply should not be very responsive to aggregate wage shocks. In fact, this intuition is confirmed with an experiment where we simulate a one period unanticipated wage change in an economy with linear wages.[31] We find that with linear wages there is no response at the extensive margin and the aggregate elasticity is quite close to the theoretical elasticity.

Our paper builds on some important recent contributions. In a model of indivisible labor, Chang and Kim (2006) show that the slope of the aggregate labor supply schedule is determined by the distribution of reservation wages rather than by the willingness to substitute leisure intertemporally, establishing that when the extensive margin is operative aggregation plays a crucial role in determining aggregate labor supply responses. More recently, Rogerson and Wallenius (2009) develop a theory of nonlinear earnings that provides some qualitative insights showing that the aggregate labor supply response to permanent tax changes is unrelated to the theoretical elasticity of labor supply implied by preference parameters. We build a closely related theory with heterogeneous agents. Our contribution is to build a theory of aggregation − disciplined with micro data − and to test the predictions of the theory for labor supply responses.

Nonlinear earnings vs. fixed costs of work. It is interesting to compare labor supply responses in our baseline economy to those of an alternative theory of the extensive margin. We thus evaluate aggregate labor supply responses in a model with linear wages in which the extensive margin is active because of fixed costs of work.[32] We find that the economy with fixed costs of work has a much lower labor supply elasticity both at the intensive and extensive margins. While the labor supply elasticity to a temporary wage change is 1.27 in the baseline economy, it is 0.59 − less than a half − in the economy with fixed costs of work. The elasticity at the extensive margin is reduced by more than a half (from 0.69 to 0.26). The response to a permanent compensated-wage change is also much smaller for the economy with fixed costs of work, especially at the extensive margin (from 0.35 to 0.18) (see Table 8). In understanding these results note that the incentives to work long hours are much weaker in the economy with fixed costs of work than in the economy with nonlinear earnings. First, in the economy with fixed costs of work individuals are less willing to work long hours because the concavity of the

[31] This economy features incomplete markets and borrowing constraints so that the main difference with the economy considered by Domeij and Flodén (2006) is that we model the life cycle.

[32] Note that in the absence of fixed costs of work, the extensive margin is not active in the presence of linear wages. To keep homotheticity, the fixed cost of work is formulated in terms of time rather than goods. Otherwise, to be consistent with the evidence the fixed costs of work would have to change across individuals in the income distribution, over time, and across countries. This alternative economy is calibrated to the targets used in the calibration of the baseline economy. We find that, in order to match the calibration target for the fraction of people working 3 quadrimesters we need the variance of the transitory wage shocks to be roughly 5 times the variance in the baseline economy with nonlinear earnings. As a result, the calibrated economy with fixed costs of work matches all the calibration targets but the variance of transitory wages, which is higher than in the data.

utility function implies that the marginal utility of leisure decreases more steeply with working hours than in the economy with no fixed costs of work. Second, contrary to the nonlinear earnings economy, the hourly wage rate does not rise with working hours in the economy with fixed costs of work. As a result, the elasticity of labor supply along the intensive margin is lower in the economy with fixed costs of work than in the economy with nonlinear earnings (0.43 versus 0.61). To understand the low labor supply response at the extensive margin, note that the more costly it is for individuals to work long hours the more costly it is for them to take periods off work.[33] The economy with fixed time costs of work exhibits low labor supply responses both along the intensive and extensive margins. The nonlinear earnings economy represents a parsimonious theory of labor supply decisions with an active extensive margin. Notice that modeling the labor market participation decision in terms of fixed costs of work requires the fixed costs to change over time if there is technological progress in the economy, or to change across individuals for the theory to be consistent with the micro facts on labor supply, or to take a stand on how the fixed costs vary across countries with different level of technology. A key advantage of our theory is that we do not need to take a stand on how fixed costs vary over time, individuals, and space (see Erosa et al. (2012)).

7 Conclusion

We build a micro-founded theory of aggregate labor supply which accounts for choices at both the intensive and the extensive margin. The key feature of our theory for delivering periods of non-participation is the nonlinear mapping between hours of work and earnings, which is convex at low hours of work. This mapping is the competitive equilibrium outcome of an economy with a production technology in which hours of work and number of workers are imperfect substitutes. The model captures salient features of labor supply over the life cycle, which is crucial for a theory of aggregation. We find that the elasticity of aggregate labor supply to a one-period wage change is 1.27, a value that is more than twice as large as the Frisch elasticity of labor supply that was hard-wired into the model through the calibration of the preference parameters. Consistent with the empirical evidence, the aggregate labor supply response is mostly driven by the extensive margin. Nonlinear earnings are crucial for a substantial response at the extensive margin and deliver a parsimonious theory of individual labor supply and aggregate responses. We test the predictions of the theory for aggregate labor supply responses with some of the quasi-experimental evidence reviewed in Chetty et

[33] This also explains why the calibration of the economy with fixed costs of work requires large transitory shocks to wages.

al. (2011). We find that the model economy is consistent with the micro-estimates of the Hicks elasticity of labor supply (steady-state tax changes). Unlike the findings reported in Chetty et al. (2011) regarding the Rogerson and Wallenius (2009) model, we find that our theory is consistent with the evidence from the Icelandic temporary tax reduction that took place in 1987. We conclude that our theory, calibrated to micro data, is consistent with a Frisch elasticity of aggregate labor supply well above 1. Nonetheless, our estimate for the employment elasticity is well below the value of 2 needed to match the business cycle data along the extensive margin. Our theory develops a neoclassical model of labor markets with life cycle and heterogeneous agents which abstracts from unemployment. This non-trivial extension is left for future research (see Bils et al. (2009) for an early step in this direction).

Table 1: The Coefficient of Variation in Lifetime and Cross-sectional Hours, Data vs. Model.

	Cross-sectional		Lifetime	
Age	Data	Model	Data	Model
26-35	0.36	0.20	0.27	0.08
36-45	0.34	0.22	0.26	0.11
46-55	0.41	0.25	0.37	0.15
56-64	0.86	0.47	0.64	0.36

Table 2: PSID: Stochastic Process of Hourly Wages.

	Non-college	College
Var(α)	0.097	0.072
ρ	0.940	0.977
Var(η)	0.019	0.021

Table 3: Calibration.

Parameter	Value	Variable	Target	Result
$\overline{W}$	3.3	Ratio of mean economy to mean male earnings	0.80	0.82
φ	1.0	Fraction of hours worked, age 30-45	42%	42%
β	0.9815	Asset to income ratio	2.5	2.6

Table 4: Calibration of the Stochastic Process for Wages.

Parameter	Value	Variable	Target	Model
		Non-College		
σ_α	0.283	Variance of fixed component of log wages	0.10	0.09
ρ	0.937	Correlation coefficient AR component of log wages	0.94	0.94
σ_η	0.138	Variance of innovation AR component of log wages	0.02	0.02
		College		
Parameter	Value	Variable	Target	Model
σ_α	0.236	Variance of fixed component of log wages	0.07	0.07
ρ	0.972	Correlation coefficient AR component of log wages	0.98	0.98
σ_η	0.117	Variance of innovation AR component of log wages	0.02	0.02

Table 5: Labor Supply, Intensive and Extensive Margin: Model vs. Data.

	Extensive Margin		Intensive Margin	
	Model	Data	Model	Data
age	-0.0066	-0.0069	-0.0017	-0.0010
	(0.0001)	(0.0005)	(0.0001)	(0.0005)
education	0.0403	0.0931	0.0154	0.0374
	(0.0012)	(0.0098)	(0.0009)	(0.0078)
assets	-0.0003	-0.00006	0.0002	0.0001
	(0.0001)	(0.00006)	(0.00002)	(0.0001)

Notes: In order to analyze the extensive margin, we use a probit regression of employment on age, education, and assets. For the analysis on the intensive margin we use an OLS regression of log hours on age, education, and assets. The analysis on the SIPP data also includes a race dummy variable.

Table 6: The Covariance Between the Change in Log Hours and the Change in Log Wages, by Age Groups.

Age	26-35	36-45	46-55	56-64
Data	-0.04	-0.03	-0.03	-0.04
$\sigma = 1.8$	-0.03	-0.01	-0.02	-0.04
$\sigma = 2.0$: Baseline	**-0.02**	**-0.01**	**-0.01**	**-0.03**
$\sigma = 2.2$	-0.01	0.00	0.00	-0.01
$\sigma = 2.0$: Fixed cost	0.01	0.03	0.02	0.01

Table 7: The Correlation Between the Change in Log Hours and the Change in Log Wages, by Age Groups.

Age	26-35	36-45	46-55	56-64
Data	-0.26	-0.27	-0.24	-0.24
$\sigma = 1.8$	-0.16	-0.11	-0.12	-0.16
$\sigma = 2.0$: Baseline	**-0.13**	**-0.06**	**-0.06**	**-0.10**
$\sigma = 2.2$	-0.02	-0.06	0.02	-0.01
$\sigma = 2.0$: Fixed cost	0.07	0.15	0.11	0.02

Table 8: Aggregate Labor Supply Elasticity.

	Temporary wage increase			Permanent wage decrease		
Elasticities:	Total	Employment	Intensive	Total	Employment	Intensive
$\sigma = 1.8$	1.56	0.91	0.65	0.71	0.43	0.28
$\sigma = \mathbf{2.0}$**: Baseline**	**1.27**	**0.69**	**0.58**	**0.65**	**0.35**	**0.30**
$\sigma = 2.2$	1.07	0.55	0.52	0.57	0.27	0.30
$\sigma = 2.0$: Fixed cost	0.59	0.26	0.33	0.45	0.18	0.27

Table 9: Aggregate Labor Supply Elasticity by Age: Baseline Economy.

	Temporary wage increase			Permanent wage decrease		
Elasticities:	Total	Employment	Intensive	Total	Employment	Intensive
26-35	1.0	0.38	0.62	0.70	0.26	0.43
36-45	1.17	0.53	0.64	0.62	0.31	0.21
46-55	1.35	0.72	0.63	0.56	0.31	0.25
56-64	1.98	1.56	0.42	0.72	0.53	0.18
All	1.27	0.69	0.58	0.65	0.35	0.29

Table 10: Aggregate Labor Supply Elasticity by Age: The Iceland Experiment.

	Non-college			College		
Elasticities:	Total	Employment	Intensive	Total	Employment	Intensive
26-35	0.50	0.20	0.31	0.46	0.17	0.29
36-45	0.59	0.24	0.35	0.55	0.16	0.38
46-55	0.74	0.37	0.38	0.61	0.20	0.41
56-64	1.21	0.85	0.36	1.12	0.70	0.43
All	0.69	0.35	0.34	0.64	0.27	0.37

Figure 1: Mean Annual Hours Worked, 1968-1996, PSID, Men, All and by Education.

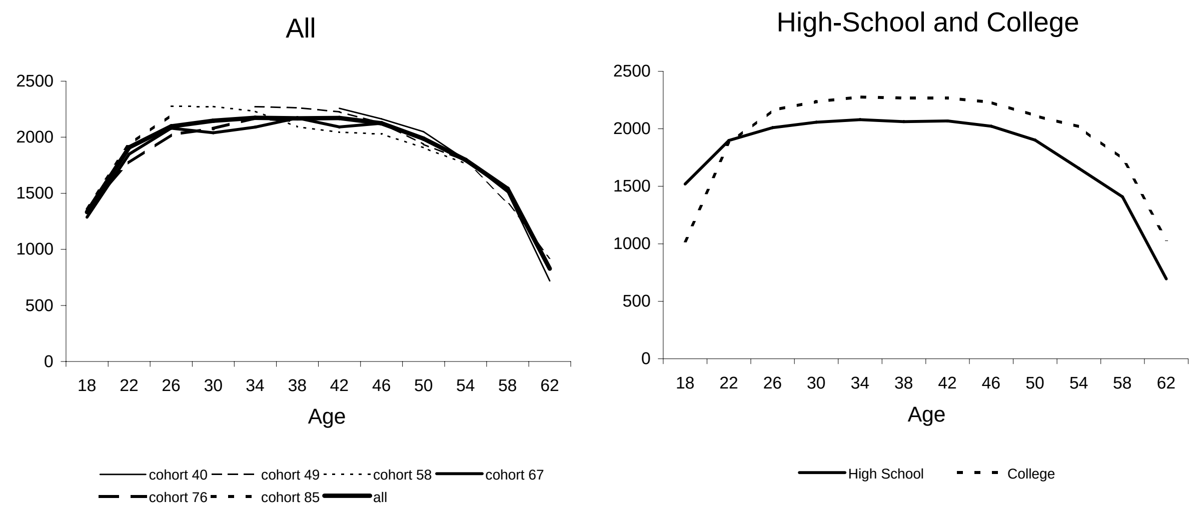


Figure 2: Coefficient of Variation of Annual Hours, 1968-1996, PSID Men, All and by Education.

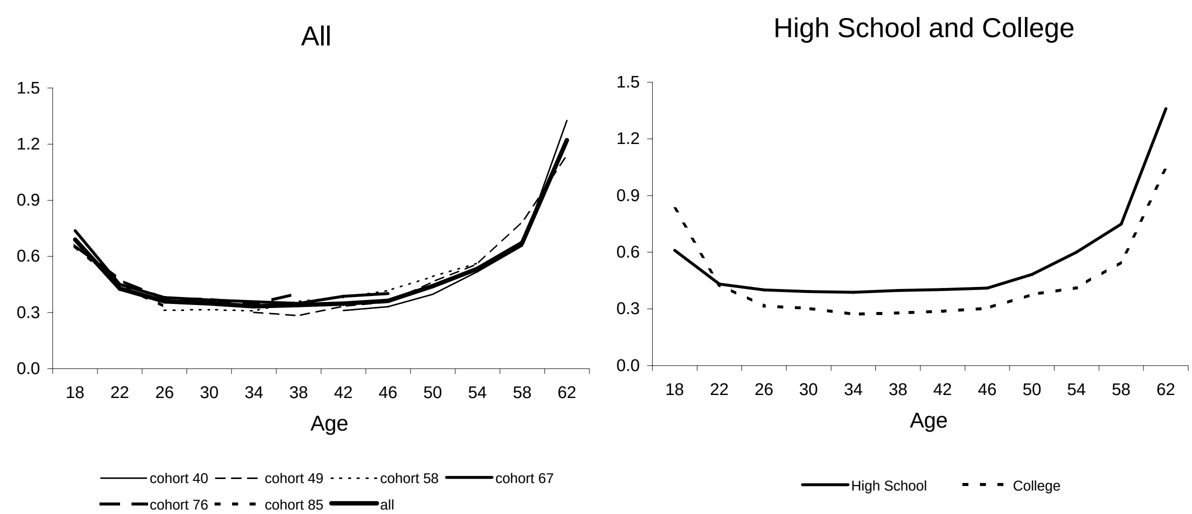


Figure 3: Persistence of Annual Hours, 1968-1996, PSID, Men.

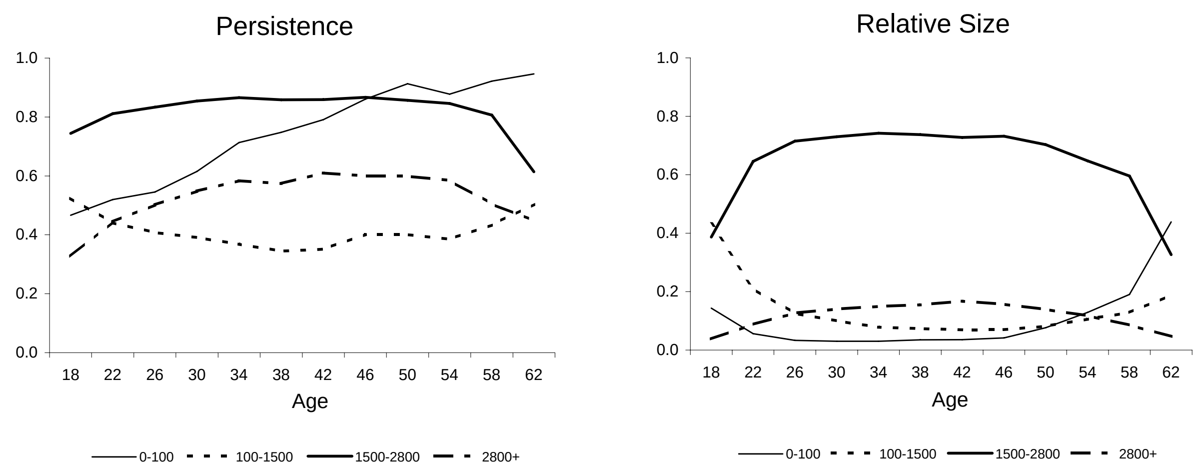


Figure 4: The Effect of Nonlinear Wages on Labor Supply in a Static Model.

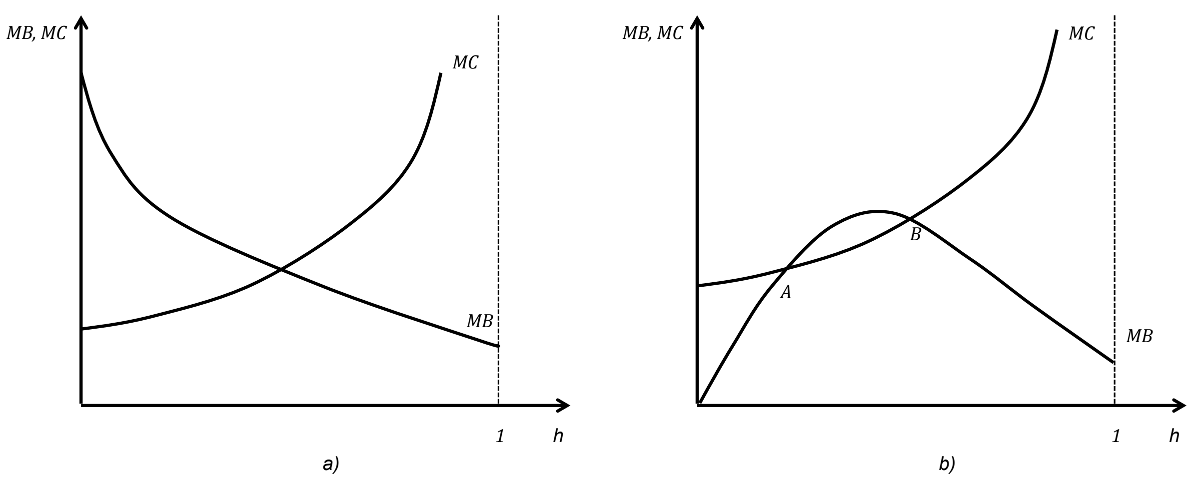
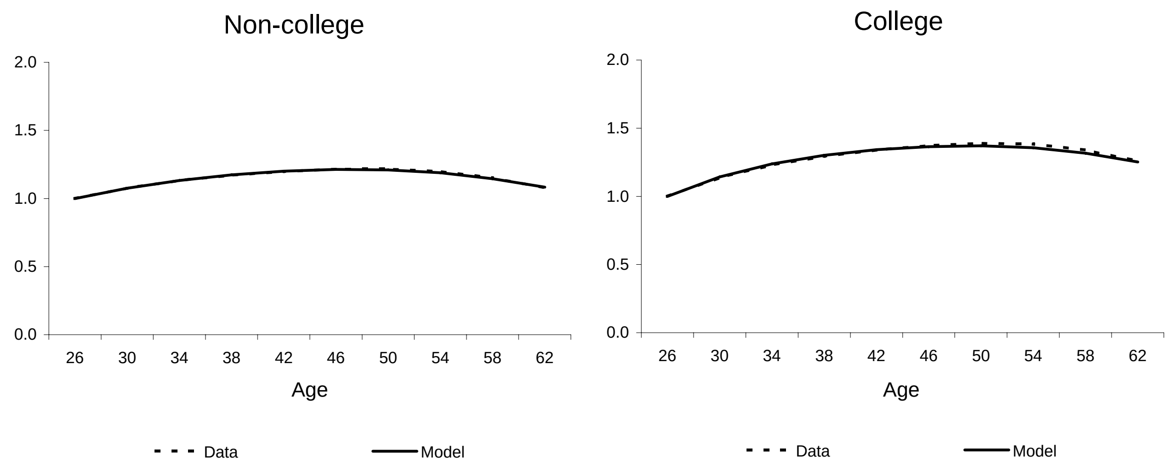


Figure 5: The Life-cycle Deterministic Component of Wages, by Education, Model vs. Data.



Note: On the graphs, by education, wages have been normalized to 1 for the first age group.

Figure 6: The Fraction of Individuals Working Three Quadrimesters in a Year, by Education, Model vs. Data.

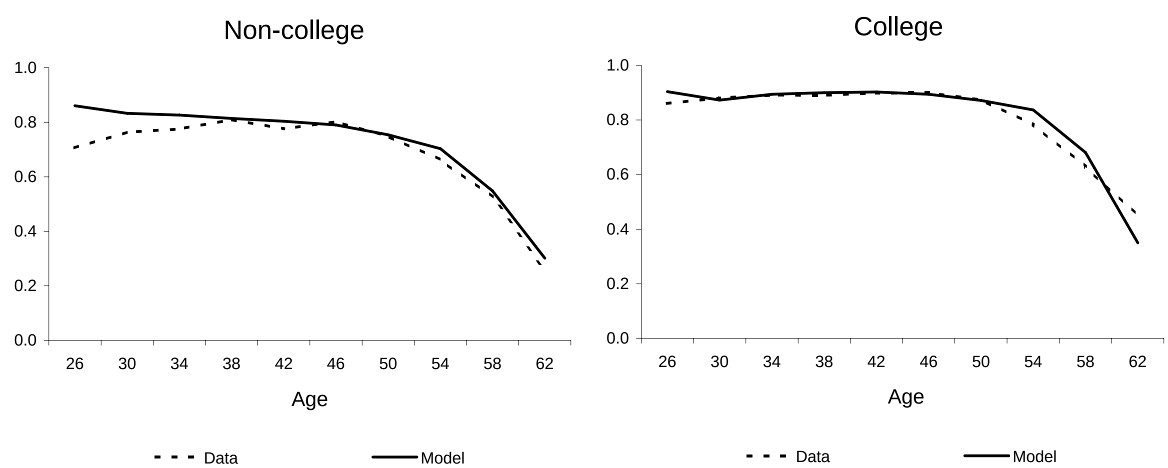


Figure 7: Variance of the Transitory Component of Residual Log Wages, by Education, Model vs. Data.

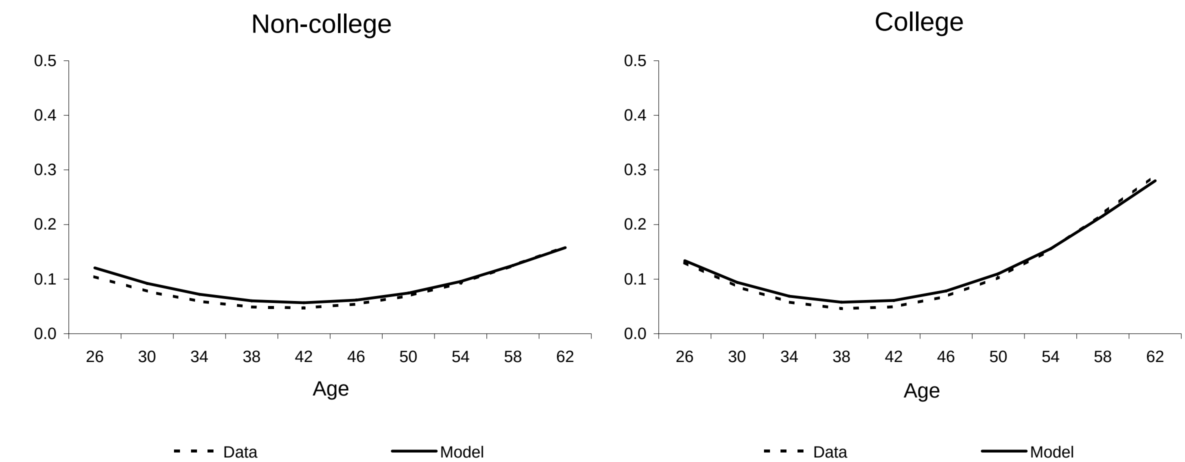


Figure 8: Mean Annual Hours Worked, Model vs. Data.

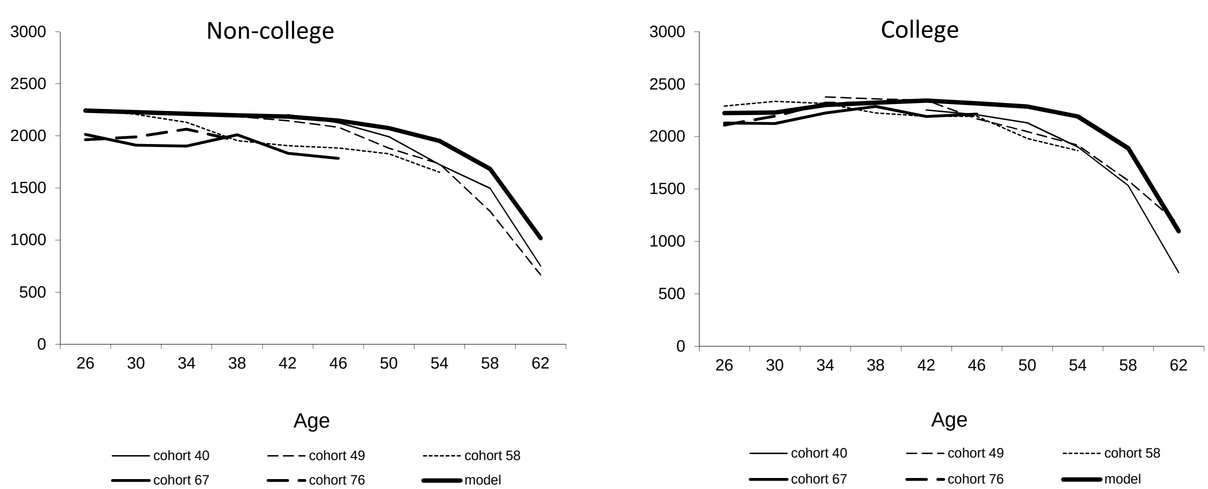


Figure 9: Mean Annual Hours Worked, Workers with Positive Hours, Model vs. Data.

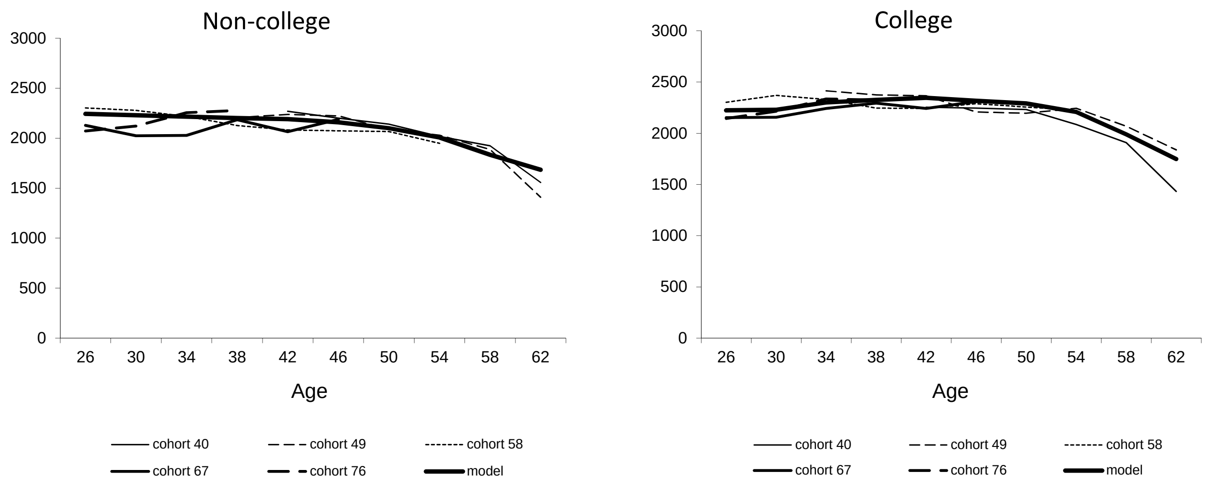


Figure 10: Coefficient of Variation of Annual Hours, Model vs. Data.

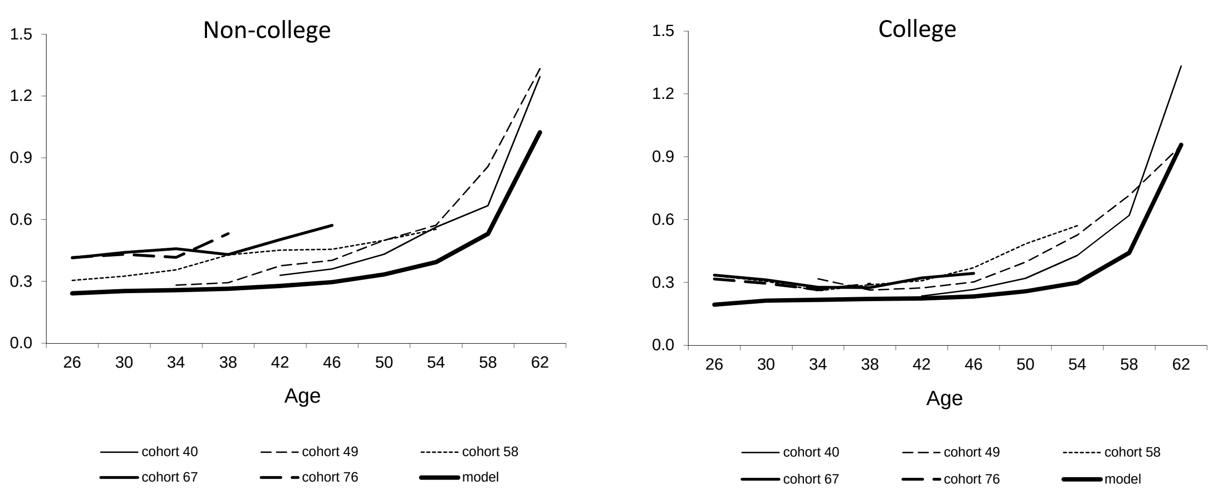


Figure 11: Annual Hours Groups, Model vs. Data.

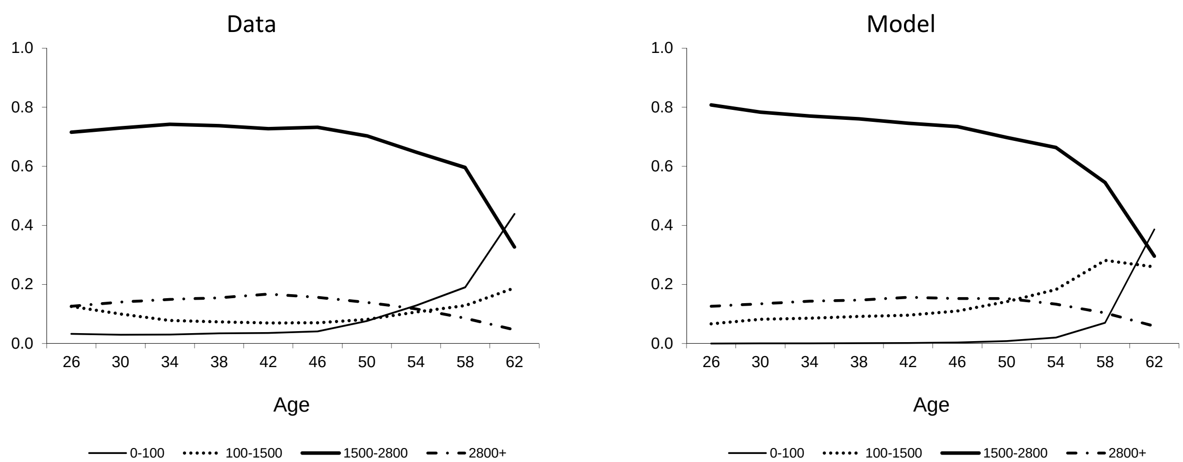


Figure 12: Persistence of Annual Hours, Model vs. Data.

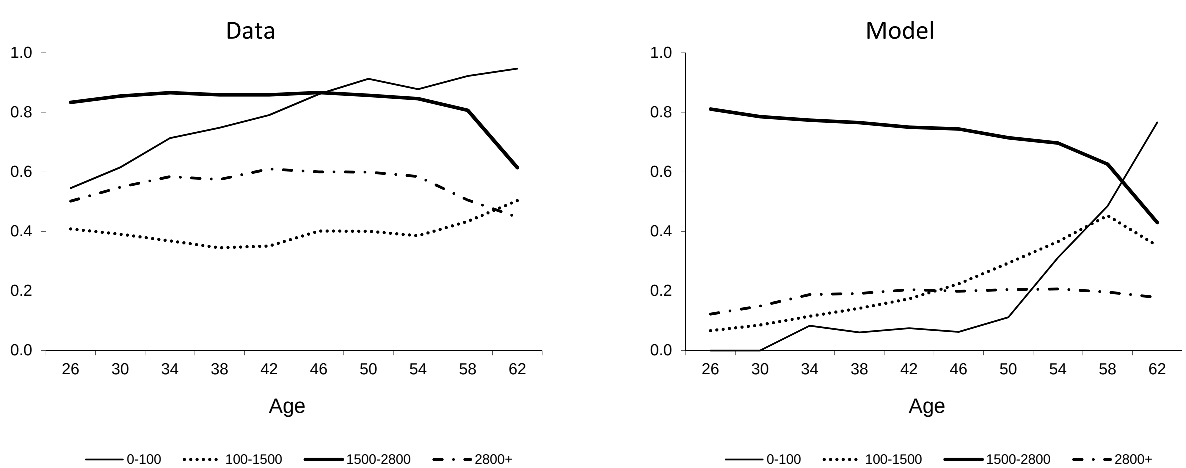
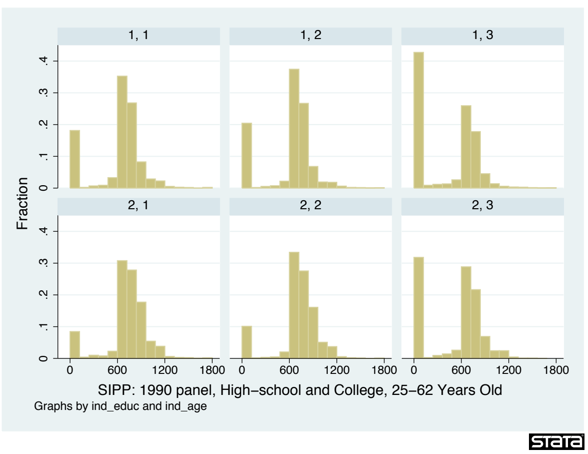
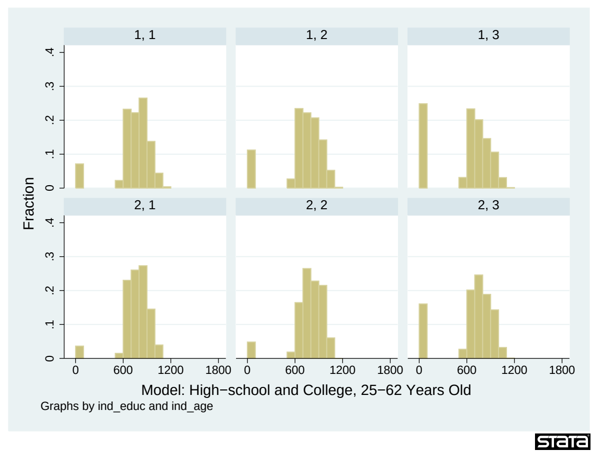


Figure 13: The Distribution of Hours within a 4-month Period, Model vs. Data.





Notes: The graphs show the distribution of hours in a 4-month period for age and education groups, in the data and in the model. Group ij represents a particular age (i) and education (j) group. i: 1. 26-40, 2. 41-55, 3. 56-61; j: 1. non-college, 2. college.

References

Aaronson, Daniel and Eric French, "The Effect of Part Time Work on Wages: Evidence from Social Security Rules," *Journal of Labor Economics*, 2004, *22* (2), 329–252.

_ **and** _ , "The Effect of Progressive Taxation on Labor Supply when Hours and Wages are Jointly Determined," *Journal of Human Resources*, Spring 2009, *44* (2), 386 – 408.

Bhattacharya, Jay and Darius Lakdawalla, "Does Medicare Benefit the Poor?," *Journal of Public Economics*, January 2006, *90* (1-2), 277–292.

Bianchi, Marco, Bjorn R. Gudmunsdsson, and Gylfi Zoega, "Iceland's Natural Experiment in Supply-Side Economics," *American Economic Review*, 2001, *91* (5), 1564–1579.

Bils, Mark, Yongsung Chang, and Sun-Bin Kim, "Comparative Advantage and Unemployment," *RCER Working Papers, University of Rochester*, 2009, *547.*

Blundell, Richard, Antoine Bozio, and Guy Laroque, "Extensive and Intensive Margins of Labour Supply: Working Hours in the US, UK and France," IFS Working Paper 01/11, Institute for Fiscal Studies 2011.

Castaneda, Ana, Javier Díaz-Giménez, and José-Víctor Ríos-Rull, "Exploring the Income Distribution Business Cycle Dynamics," *Journal of Monetary Economics*, June 1998, *42* (1), 93–130.

Chang, Yongsung and Sun-Bin Kim, "From Individual to Aggregate Labor Supply: A Quantitative Analysis Based on a Heterogeneous Agent Macroeconomy," *International Economic Review*, February 2006, *47* (1), 1–27.

Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber, "Does Indivisible Labor Explain the Difference Between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities," Working Paper No. 16279, NBER 2011.

Conesa, Juan Carlos, Sagiri Kitao, and Dirk Krueger, "Taxing Capital? Not a Bad Idea after All," *American Economic Review*, 2009, *99* (1), 25–48.

Cooley, Thomas F., ed., *Frontiers of Business Cycle Research*, Princeton: Princeton University Press, 1995.

Domeij, David and Jonathan Heathcote, "On the Distributional Effects of Reducing Capital Taxes," *International Economic Review*, 2004, *45* (2), 523–554.

— **and Martin Flodén**, "The Labor-Supply Elasticity and Borrowing Constraints: Why Estimates are Biased," *Review of Economic Dynamics*, 2006, *9*, 242–262.

Erosa, Andrés, Luisa Fuster, and Gueorgui Kambourov, "Labor Supply and Government Programs: A Cross-Country Analysis," *Journal of Monetary Economics*, 2012, *forthcoming.*

French, Eric, "The Effects of Health, Wealth, and Wages on Labor Supply and Retirement Behavior," *Review of Economic Studies*, April 2005, *72* (2), 395–427.

Gourieroux, C., A. Monfort, and E. Renault, "Indirect Inference," *Journal of Applied Econometrics*, December 1993, *8* (S), S85–112.

Gruber, Jonathan and David A. Wise, eds, *Social Security Programs and Retirement around the World*, Chicago: University of Chicago Press, 1999.

Guvenen, Fatih and Anthony Smith, "Inferring Labor Income Risk from Economic Choices: An Indirect Inference Approach," *Staff Report Federal Reserve Bank of Minneapolis*, 2010, *450.*

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante, "The Macroeconomic Implications of Rising Wage Inequality in the United States," *Journal of Political Economy*, August 2010, *118* (4), 680–722.

Heckman, James J. and Thomas MaCurdy, "A Life-Cycle Model of Female Labor Supply," *The Review of Economic Studies*, 1980, *47* (1), 47–74.

Hornstein, Andreas and Edward C. Prescott, "The Firm and the Plant in General Equilibrium Theory," in Robert Becker, Michele Boldrin, Ronald Jones, and William Thomson, eds., *General Equilibrium, Growth, and Trade II: The Legacy of Lionel McKenzie*, San Diego: Academic Press, 1993.

Imai, Susumu and Michael P. Keane, "Intertemporal Labor Supply and Human Capital Accumulation," *International Economic Review*, 2004, *45* (2), 601–641.

Kaplan, Greg, "Inequality and the Lifecycle," mimeo, University of Pennsylvania 2011.

— **and Giovanni L. Violante**, "How Much Consumption Insurance Beyond Self-Insurance?," *American Economic Journal: Macroeconomics*, October 2010, *2* (4), 53–87.

Keane, Michael P., "Labor Supply and Taxes: A Survey," Working Paper No. 160, UTS School of Finance 2010.

_ **and Richard Rogerson**, "Reconciling Micro and Macro Labor Supply Elasticities: A Structural Perspective," Working Paper 17430, NBER 2011.

Kimmel, Jean and Thomas J. Kniesner, "New Evidence on Labor Supply: Employment versus Hours Elasticities by Sex and Marital Status," *Journal of Monetary Economics*, 1998, *42*, 289–301.

Ljungqvist, Lars and Thomas J. Sargent, "Career Length: Effects of Curvature of Earnings Proles, Earnings Shocks, Taxes, and Social Security," mimeo, New York University 2011.

Low, Hamish, "Self-insurance in a Life-cycle Model of Labour Supply and Savings," *Review of Economic Dynamics*, 2005, *8* (4), 945–975.

Osuna, Victoria and José-Víctor Ríos-Rull, "Implementing the 35 Hour Workweek by Means of Overtime Taxation," *Review of Economic Dynamics*, 2003, *6* (1), 179–206.

Pijoan-Mas, Josep, "Precautionary Savings or Working Longer Hours?," *Review of Economic Dynamics*, 2006, *9*, 326–352.

Prescott, Edward, "Why do Americans Work much more than Europeans?," *Quarterly Review Federal Reserve Bank of Minneapolis*, 2004, *July*, 2–13.

Rogerson, Richard, "Indivisible Labor, Lotteries and Equilibrium," *Journal of Monetary Economics*, 1988, *21*, 3–16.

_ **and Johanna Wallenius**, "Micro and Macro Elasticities in a Life Cycle Model with Taxes," *Journal of Economic Theory*, 2009, *144*, 2277–2292.

Smith, Anthony, "Three Essays on the Solution and Estimation of Dynamic Macroeconomic Models," *PhD Dissertation, Duke University*, 1990.

Storesletten, Kjetil, Chris Telmer, and Amir Yaron, "How Important Are Idiosyncratic Shocks? Evidence from Labor Supply," *American Economic Review Papers and Proceedings*, 2001, *91*, 413–417.

Tauchen, George, "Finite State Markov-Chain Approximations to Univariate and Vector Autoregressions," *Economics Letters*, 1986, *20* (2), 177–181.

Wallenius, Johanna, "Human Capital and the Intertemporal Elasticity of Substitution of Labor," mimeo, Arizona State University 2009.

APPENDIX

The empirical elasticity of labor supply. The empirical elasticity of labor supply is obtained by estimating the following regression on model generated data via ordinary least squares (OLS) and instrumental variables (IV):

$$\triangle \ln h_{it} = \beta_0 + \beta_1 \triangle \ln w_{it} + \varepsilon_{it}, \tag{A-1}$$

where the regression coefficient β_1 gives the empirical labor supply elasticity predicted by the model economy. Since the error term may be correlated with the contemporaneous wage change due to the wealth effect of a wage change on labor supply, it is standard to follow an instrumental variables approach. The instruments used are past wage changes (IV1) and a composite of a constant, age, age-squared, and the twice lagged log-wage (IV2). The findings are reported in Table A-1.

We find that all the estimates of the empirical elasticity of labor supply are well below the value predicted by the theoretical elasticity. The lowest estimates are obtained when using OLS and when instrumenting with lagged wage changes (IV1). The highest estimates are obtained when instrumenting with a constant, age, age-squared, and the twice lagged log-wage (IV2). Note that the IV2 instruments use information from the age-profile of wages and hours to estimate the labor supply elasticity. This procedure identifies the elasticity from anticipated life-cycle wage changes and is particularly good in correcting for measurement error in hours and wages. In the baseline economy, the empirical elasticity of labor supply obtained with IV2 is 0.36, well within the range of $[0, 0.5]$ in the empirical literature. Interestingly, the theoretical elasticity is equal to 0.61, almost twice the value implied by the empirical elasticity.

Table A-1: The Elasticity of Labor Supply.

	Theoretical	OLS	IV1 ($\Delta \ln w_{t-1}$)	IV2 (age,age^2,$\ln w_{t-2}$)
$\sigma = 1.8$	0.68	-0.11	-0.15	0.42
$\sigma =$ **2.0: Baseline**	**0.61**	**-0.07**	**-0.08**	**0.36**
$\sigma = 2.2$	0.56	-0.01	-0.01	0.26
$\sigma = 2.0$: Fixed cost	0.44	0.09	0.13	-0.07

Discussion on time aggregation, the extensive margin, and the low empirical elasticity. We have shown that the extensive margin plays a crucial role in generating the high value of the aggregate elasticity of labor supply relative to the empirical elasticity estimated with conventional econometric techniques. We now assume that the econometrician observes *all* the micro data in our model economy with no measurement error (such as labor productivity for those individuals who do not work). The question is then: Can the econometrician recover the aggregate elasticity of labor supply in the model economy from the simulated micro data?

The wage rate in annual survey data is measured as $\ln(w) = \ln \frac{AnnualEarnings}{AnnualHours}$. We now argue that (even in the absence of measurement error in hours and earnings) this wage rate gives a noisy measure of the returns to work faced by individuals which biases estimates of the empirical elasticity of labor obtained with annual data. We run a regression (A-1) with annual model data assuming away measurement error and using the quarterly sum of realized labor productivity as an explanatory variable ($\ln \sum z$ instead of $\ln w$). In order to incorporate labor supply responses along the extensive margin we run regression (A-1) on changes in log-leisure rather than changes in log hours (to avoid the log-zero problem when individuals do not work). Table A-2 shows that the annual elasticity of leisure increases from -0.37 to -0.52 when using the quarterly sum of labor-productivity instead of annual wages. Next, we show that there are some other important subtleties in aggregating the returns to work over the year. To this end, we run the regression (A-1) using $\sum(\ln(z))$ as a regressor. While the empirical elasticity of leisure is -0.52 when using $\ln \sum z$, it is -0.57 when using $\sum(\ln(z))$. Hence, the log of the sum of labor productivities ($\ln \sum z$) is a worse measure of the returns to work faced by individuals during the year than the sum of the log of quarterly labor productivities $\sum(\ln(z))$. Because there may be important wealth effects associated with change in labor productivity over the year, we isolate the substitution effect of a change in labor productivity by instrumenting he regression (A-1) with changes in lagged $\sum(\ln(z))$. This procedure gives an empirical elasticity of leisure equal to -0.88, which fully recovers the aggregate elasticity of leisure obtained in the temporary wage experiment.

Table A-2: Elasticity of Leisure: Time Aggregation, No Measurement Error.

	Theoretical	Annual	$Ln(\sum z)$	$\sum(ln(z))$	IV $\sum(ln(z))$	Macro
$\sigma = 1.8$	-0.56	-0.40	-0.53	-0.59	-1.06	-1.01
$\sigma =$ **2.0: Baseline**	**-0.50**	**-0.37**	**-0.52**	**-0.57**	**-0.88**	**-0.85**
$\sigma = 2.2$	-0.45	-0.35	-0.50	-0.55	-0.76	-0.72
$\sigma = 2.0$: Fixed cost	-0.50	-0.27	-0.34	-0.39	-0.46	-0.43

Table A-3: Transition Matrix across Annual Hours Cells, in Percent, PSID, 1968-1996, Men.

	Ages 18-29				
		To			Relative
From	000-100	100-1500	1500-2800	2800+	Size
0-100	50.11	32.20	16.52	1.17	6.42
100-1500	7.55	46.68	42.53	3.24	22.00
1500-2800	0.81	9.73	81.34	8.11	62.19
2800+	0.32	4.15	48.78	46.75	9.39

	Ages 30-54				
		To			Relative
From	0-100	100-1500	1500-2800	2800+	Size
0-100	78.46	14.72	6.16	0.66	15.12
100-1500	10.56	37.78	47.20	4.46	8.01
1500-2800	0.60	5.48	86.02	7.90	72.86
2800+	0.26	2.61	38.78	58.35	15.12

	Ages 55-65				
		To			Relative
From	0-100	100-1500	1500-2800	2800+	Size
0-100	92.18	6.40	1.23	0.19	21.19
100-1500	31.88	42.47	23.81	1.85	13.92
1500-2800	2.22	12.80	79.44	5.54	56.17
2800+	1.37	4.61	40.52	53.51	8.72

Note: Authors' calculations from the PSID.

Table A-4: Transition Probability across Annual Hours Cells, PSID, 1968-1996, Men, High-School.

	Ages 18-29				
	To				Relative
From	000-100	100-1500	1500-2800	2800+	Size
0-100	55.03	24.12	20.56	0.29	6.90
100-1500	11.10	31.66	54.92	2.33	12.50
1500-2800	1.05	7.04	87.26	4.64	74.12
2800+	0.26	2.14	48.14	49.47	6.47
	Ages 30-54				
	To				Relative
From	0-100	100-1500	1500-2800	2800+	Size
0-100	81.37	9.85	8.43	0.36	6.26
100-1500	17.83	30.20	50.03	1.94	5.40
1500-2800	1.18	3.69	90.19	4.94	78.83
2800+	0.24	2.07	42.72	54.97	9.51
	Ages 55-65				
	To				Relative
From	0-100	100-1500	1500-2800	2800+	Size
0-100	92.35	5.30	2.35	0.0	25.55
100-1500	42.23	31.75	25.42	0.60	9.88
1500-2800	4.26	9.33	83.72	2.68	59.94
2800+	2.59	3.73	44.33	49.35	4.62

Note: Authors' calculations from the PSID.

Table A-5: Transition Probability across Annual Hours Cells, PSID, 1968-1996, Men, College.

	Ages 18-29				
	To				Relative
From	000-100	100-1500	1500-2800	2800+	Size
0-100	48.75	29.99	20.14	1.13	5.65
100-1500	7.99	42.97	47.47	1.57	15.78
1500-2800	0.65	5.77	88.08	5.49	72.64
2800+	0.38	2.07	54.50	43.05	5.93

	Ages 30-54				
	To				Relative
From	0-100	100-1500	1500-2800	2800+	Size
0-100	68.47	13.82	16.90	0.81	1.65
100-1500	8.31	31.72	55.52	4.44	3.27
1500-2800	0.44	2.28	91.42	5.86	84.55
2800+	0.24	0.98	49.03	49.76	10.54

	Ages 55-65				
	To				Relative
From	0-100	100-1500	1500-2800	2800+	Size
0-100	93.69	4.48	1.83	0.0	14.13
100-1500	30.56	37.05	29.42	2.97	21.30
1500-2800	2.32	7.70	85.90	4.08	73.00
2800+	0.0	2.94	62.06	35.0	5.70

Note: Authors' calculations from the PSID.

Table A-6: Labor Supply, by Education, Intensive and Extensive Margin: Model vs. Data.

	Extensive Margin				Intensive Margin			
	Non-College		College		Non-College		College	
	Model	Data	Model	Data	Model	Data	Model	Data
age	-0.0070	-0.0087	-0.0059	-0.0054	-0.0019	-0.0014	-0.0011	-0.0006
	(0.0001)	(0.0007)	(0.0001)	(0.0001)	(0.0001)	(0.0006)	(0.0001)	(0.0007)
assets	-0.0006	0.0003	-0.0001	-0.00005	-0.0002	0.0001	0.0002	0.0001
	(0.0001)	(0.0004)	(0.00003)	(0.00005)	(0.00005)	(0.0003)	(0.00003)	(0.0001)

Figure A-1: Mean Annual Hours Worked, Workers with Positive Hours, 1968-1996, PSID Men, All and by Education.

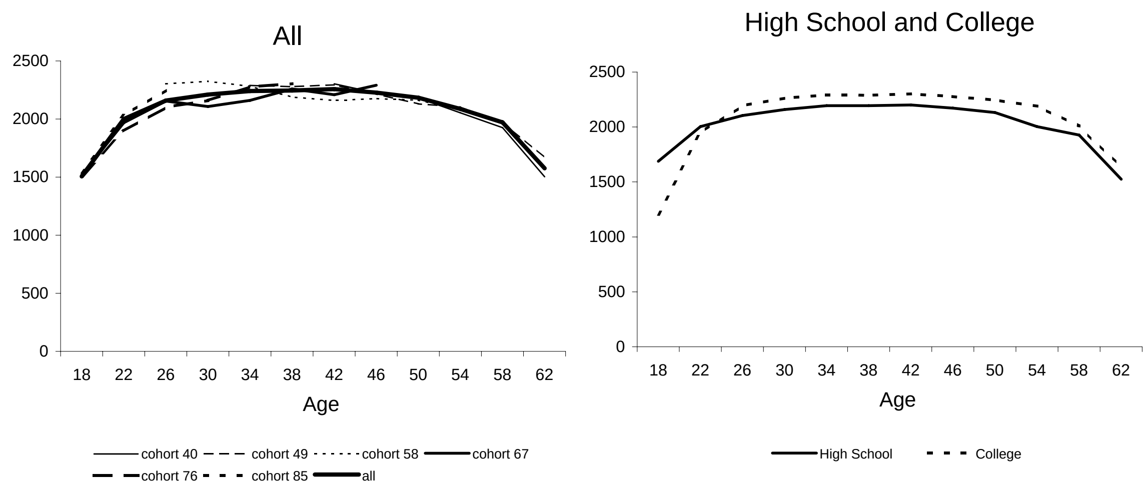


Figure A-2: Participation Rate, Fraction of Workers with Positive Annual Hours, 1968-1996, PSID Men, All and by Education.

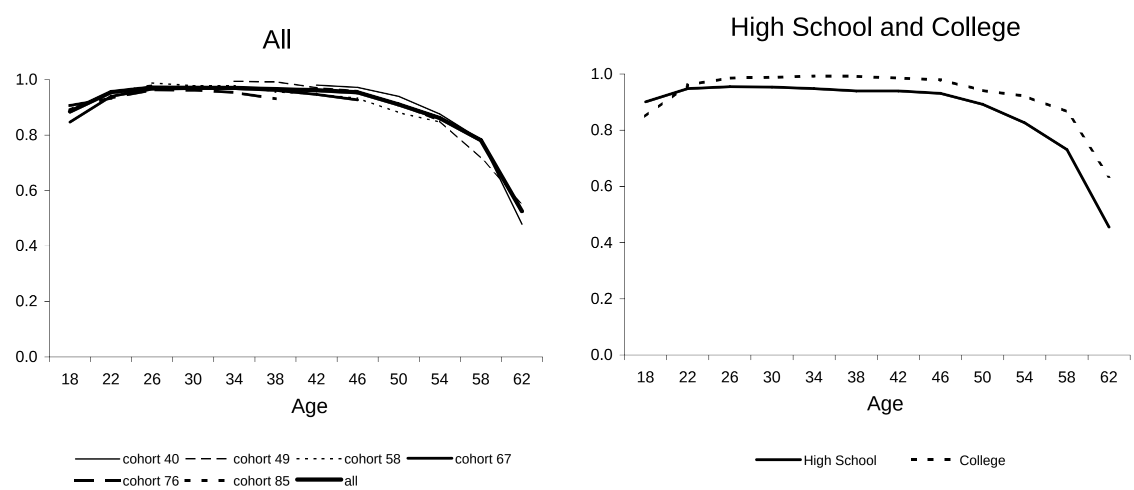
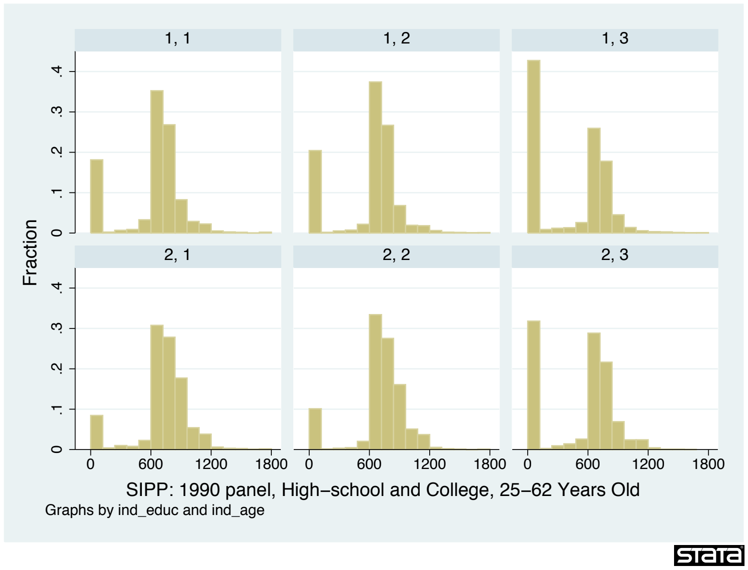


Figure A-3: The Distribution of Hours within a 4-month Period, SIPP: 1990 Panel.



Notes: The graphs show the distribution of hours in a 4-month period for age and education groups. Group ij represents a particular age (i) and education (j) group. i: 1. 26-40, 2. 41-55, 3. 56-61; j: 1. non-college, 2. college.

Figure A-4: The Fraction of Individuals Working Three Quadrimesters in a Year, by Education, SIPP: 1990 Panel.

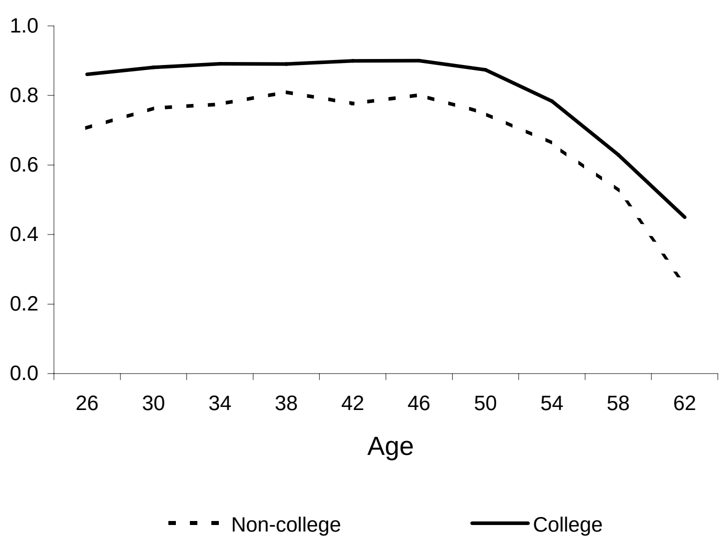


Figure A-5: Variance of the Transitory Component of Residual Log Wages, Net of the Effect of Hours Worked, by Education, Model vs. Data.

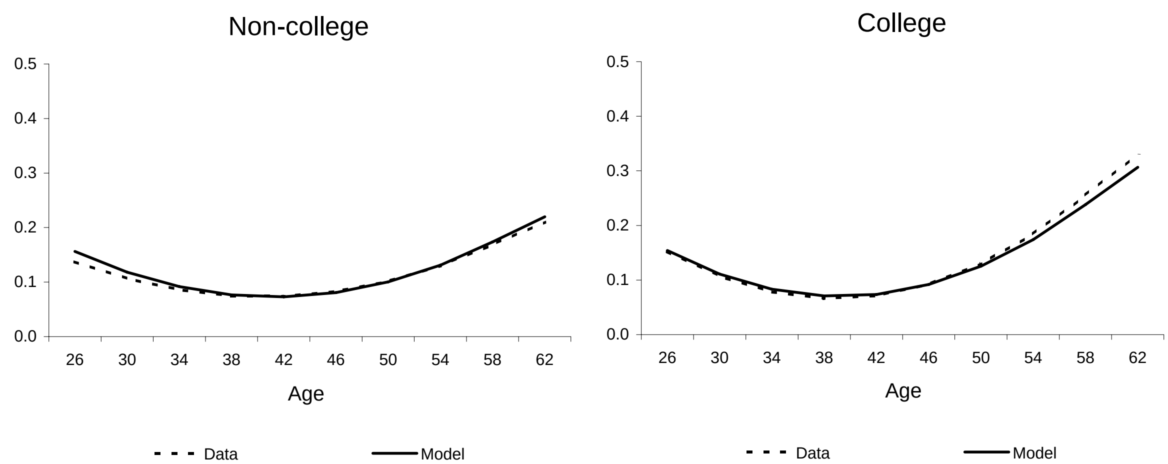


Figure A-6: The Variance of Measurement Error in Hours Worked, by Education, Model vs. Data.