The Marginal Worker and the Aggregate Elasticity of Labor Supply*

François Gourio† Pierre-Alexandre Noual‡

April 2009

Abstract

When labor supply is indivisible and markets are complete, the aggregate Frisch elasticity of labor supply depends on the shape of the distribution of reservation wages. Even if most workers are wage-inelastic, the aggregate elasticity can be large if sufficiently many agents are close to their reservation wage. To evaluate this hypothesis, we estimate the model using monthly panel data drawn from the NLSY. This allows us to measure the aggregate elasticity implied by realistic heterogeneity. We estimate that the Frisch elasticity is 1.3. This elasticity is countercyclical. Our model has a natural cross-sectional implication: workers who are nearly indifferent between working or not are more sensitive to aggregate fluctuations. We find empirical support for this prediction: on average, the group of marginal workers, which makes up 22% of the population, accounts for 49% of aggregate fluctuations in employment.

Keywords: indivisible labor, reservation wage distribution, labor supply, business cycles.

JEL Codes: E24, E32, J21, J22.

*We are grateful to seminar participants at the Federal Reserve Banks of Cleveland, Chicago, Kansas City and Philadelphia, the University of Chicago and Boston University, and at the 2006 NASM ES meetings, the 2006 SED meetings and the 2007 AEA meetings. We thank Flavio Cunha and Jing Jing Hsee for help with the data. We thank Yongsung Chang for his discussion, and we also thank Fernando Alvarez, Jeff Campbell, Claudia Olivetti, Ivan Fernandez-Val, and Lars Hansen for discussions or comments. All errors are ours.

†Corresponding author. Department of Economics, Boston University, 270 Bay State Road, Boston MA 02215. E-mail: fgourio@bu.edu.

‡Deparment of Economics, University of Chicago, and Citi. E-mail: p.a.noual@gmail.com.

1 Introduction

Researchers disagree on the value of the aggregate elasticity of labor supply. On the one hand, there is a wide consensus among labor economists that the intertemporal elasticity of hours worked per worker is low. MaCurdy (1981) found an elasticity between 0.1 and 0.4 for men who are continuously working. Further research has confirmed this finding, while suggesting that the elasticity is somewhat higher for women. By contrast, macroeconomic models typically use higher elasticities. For instance, King and Rebelo (1999) in their survey of RBC models use an elasticity of 4 in their basic model and an infinite elasticity in an extension of their model. New Keynesian models also require a high elasticity: Rotemberg and Woodford (1997; see also Woodford 2003, p. 341) use an elasticity close to 9. Any macroeconomic model which is to generate variation in aggregate labor with a wage-taking household needs to have an elastic labor supply.[1]

This discrepancy between microeconomic estimates and the macroeconomic parameters can be explained by the importance of the extensive margin. As noted by Coleman (1984), short-term fluctuations in aggregate hours worked are mostly accounted for by changes in employment, rather than by changes in hours per worker. This explains why some microeconomic estimates, such as MaCurdy's, are not directly relevant for macroeconomics. When labor supply is indivisible, most agents either work or not, irrespective of the level of the aggregate wage, and only a few agents actually shift in and out of the workforce in response to a change in the aggregate wage. One can imagine that a small group of agents ("marginal workers") does all the cyclical adjustment in the workforce, generating what looks like a highly elastic *aggregate* labor supply, while the majority of the workforce is irresponsive to macroeconomic conditions.

In this paper, we study the implications of indivisible labor when agents are heterogeneous in tastes and abilities. Our model provides an explicit link between micro data and the aggregate elasticity that is relevant for a macroeconomist, i.e. the number that a researcher needs to 'plug' into his favorite business cycle model. We show that the aggregate Frisch elasticity of labor supply depends on the homogeneity of the workforce at the margin: the aggregate elasticity

[1] Search and matching models of the labor market similarly fail to generate enough employment volatility (Shimer 2005), except when the implicit elasticity of labor supply is high (e.g. Hagedorn and Manovskii 2005, Bils Chang and Kim 2007, and the discussion in Hall 2006).

is the hazard rate of the distribution of reservation wages, evaluated for the worker who is indifferent between working and not working this period. This formula implies that the *shape* of the distribution of reservation wages around the marginal worker is the crucial determinant of the elasticity. In Rogerson's (1988) seminal paper, all agents are indifferent to working, so that the hazard rate is very high (actually infinite), and a small change in the wage is enough to induce these agents to enter the workforce, making the aggregate labor supply very (infinitely) elastic.[2] Hence when workers are very homogenous at the margin, the elasticity is very high. The key question is: *what is the elasticity implied by realistic heterogeneity, i.e. by a realistic distribution of reservation wages*?

To answer this question, we directly estimate the complete market model on panel data from the NLSY. We allow for non-parametric heterogeneity by using fixed effects that control for permanent differences in tastes, abilities, or wealth. We use monthly data for which the indivisibility is a good approximation. We find a Frisch elasticity of labor supply around 1.5. This is an average across populations with different elasticities. These populations can first be defined in terms of observable characteristics: for instance, young people, women, and less-educated workers are more elastic.[3] But there are also individual unobservable characteristics which make some agents more elastic. We consider an interesting cross-sectional implication of our model: agents which are nearly indifferent between working and not - the "marginal workers" - are more sensitive to aggregate fluctuations. These agents are indeed likely to enter or exit the labor market upon a small change in their wage. We find empirical support for this prediction in the NLSY: on average, the group of marginal workers, which makes up 22% of the population, accounts for 49% of aggregate fluctuations in employment. (In a previous draft, we obtained similar results using the PSID; see Appendix D for a brief discussion.)

Interestingly, we also find that the labor supply elasticity is countercyclical. This implies that the response of the economy to shocks is state-dependent. Moreover, an econometrician observing data generated from an economy where the labor supply elasticity is time-varying would infer that there is "wedge" in the household first-order condition for labor. This is a

[2] Following Cho (1995) and Mulligan (2001), we note that Rogerson's result does not rely on the availability of lotteries.

[3] As noted by many authors including Clark and Summers (1981), Kydland (1984), Heckman (1993), Gomme, Rogerson, Rupert and Wright (2004), Jaimovich and Siu (2007).

data puzzle that several researchers (e.g. Hall 1997) have noted before. We find however, that the time-varying elasticity does not seem to create by itself a wedge that is as volatile and procyclical as it is in the data.

Related Literature

Our paper is related to the large macroeconomic literature on the extensive margin of labor supply (e.g. Rogerson 1988, Cho and Rogerson 1988, Cho 1995, Mulligan 2001, Chang and Kim 2006 and 2008, Haefke and Reiter 2006, Rogerson and Wallenius 2007). Some of our theoretical results are similar to Mulligan (2001), but we draw different empirical implications from these results. Our work is also closely related to Chang and Kim (2006). Our work complements theirs by emphasizing the importance of distributions, while they emphasize the role of incomplete markets. This leads us to a different empirical strategy: complete markets simplify our derivations and allow us to study more clearly the effect of the shape of the distribution, which we demonstrate in some simple examples. Moreover, we can be much less restrictive on the shape of heterogeneity.[4]

A point of semantics: in this paper, when we refer to an elasticity, it is always the Frisch (or intertemporal) elasticity, i.e. the response of aggregate hours to an increase in the wage, holding marginal utility of wealth constant. This elasticity is the relevant one for business cycle analysis.

The rest of the paper is organized as follows. Section 2 develops our model of indivisible labor supply with heterogeneity and complete markets. Section 3 discusses the estimation strategy. Section 4 presents our empirical results, both for the aggregate labor supply and for the cross-sectional implications. Section 5 conducts some robustness exercises, and section 6 concludes.

2 A complete markets model of indivisible labor supply

In this section we analyze a model of indivisible labor supply with complete markets and heterogeneity. There are three main results. First we derive an equation for labor force par-

[4] See also Imai and Keane (2004) for an analysis of the intensive margin of labor supply which incorporates human capital accumulation.

ticipation, which we estimate in Section 4. Second, we show that the aggregate labor supply curve is determined by the distribution of agents' characteristics (abilities, tastes and wealth), i.e. by the distribution of reservation wages. As a result, the aggregate Frisch elasticity of labor supply is the hazard rate of this distribution, evaluated at the marginal worker. Finally, we derive a cross-sectional prediction: some workers are more sensitive to aggregate shocks.[5]

A. **Economic Environment**

Time is discrete, indexed by $t = \{0, 1, 2, ...\}$. There is a unit measure of agents, indexed by $i \in [0, 1]$. Labor is indivisible: each agent works either $\overline{n}$ hours or not at all. We abstract from the intensive margin because it is relatively unimportant for business cycles and because microeconomic estimates of the elasticity of hours are low.

We assume separable preferences: working does not affect the utility $u(c_{it})$ derived from consumption. Individual preferences are characterized by the function $u(\cdot)$ and the disutility of work $v(\overline{n})$. We normalize $v(0) = 0$.

We denote agent i's productivity by $\pi_{it} \in \mathbb{R}^{++}$; π_{it} is a stochastic endowment of efficiency units of labor. Agent i has a stochastic disutility for labor $\theta_{it} \in \mathbb{R}^{++}$. At each date t, agents are subject to two shocks: an idiosyncratic shock to their productivity and an idiosyncratic shock to their tastes. These shocks are potentially correlated. The state of agent i at date t is summarized by the vector $s_{it} = (\pi_{it}, \theta_{it}) \in S = (\mathbb{R}^{++})^2$. We assume that s_{it} follows a stationary Markov process with a unique invariant distribution. This process implies a density p_s^t over histories: $p_s^t(s^t) = p_s^t(\pi^t, \theta^t)$ is the probability of history $s^t = (s_1, ..., s_t)$. By the law of large numbers, $p_s^t(s^t)$ is also the fraction of people with history s^t.

To sum up, preferences are:

$$U_i = E \sum_{t \geq 0} \beta^t \left(u(c_{it}) - \theta_{it} v(n_{it}) \right).$$

Because our paper deals only with labor supply, we do not need to specify the technology and the resource constraints: we only derive the aggregate "labor supply curve". More precisely, in the next section we show that this economy has a well-defined aggregate utility function and we characterize it. It is straightforward to close this model with a standard neoclassical

[5] Some of our results are closely related to Mulligan (2001).

production and resource constraint, and given some exogenous shocks to compute the dynamic equilibrium of the economy.

B. **Participation Equation**

We consider the problem of a hypothetical social planner, who maximizes the weighted sum of utilities of households, subject to the constraints that he provides a certain aggregate labor process $\{N_t\}$ to "the market' (where N_t is measured in efficiency units of labor) and that he uses a given aggregate consumption process $\{C_t\}$. Denote by z_t the aggregate state which governs these processes, and let $p_z^t(z^t)$ be the probability of history $(z_1, ..., z_t)$. The Pareto weights are $\mu_i \geq 0$ and we denote their (arbitrary) distribution by H. Given $\{C_t(z^t), N_t(z^t)\}$, the planner's objective is to maximize by choice of $\left\{c_{it}(\pi^t, \theta^t, z^t), n_{it}(\pi^t, \theta^t, z^t)\right\}$:

$$\int_0^1 \mu_i U_i di = \sum_{t \geq 0} \sum_{z^t \in Z^t} \beta^t p_z^t(z^t) \int_0^1 \int_{\theta^t} \int_{\pi^t} \mu_i \left[u(c_{it}(\pi^t, \theta^t, z^t)) - \theta_t v(n_{it}(\pi^t, \theta^t, z^t))\right] p_s^t(\pi^t, \theta^t) d\pi^t d\theta^t di,$$

subject to the constraints:

$$\forall t \geq 0, \forall z^t \in Z^t : \int_0^1 \int_{\theta^t} \int_{\pi^t} \pi_t n_{it}(\pi^t, \theta^t, z^t) p_s^t(\pi^t, \theta^t) d\pi^t d\theta^t di = N_t(z^t),$$
$$\forall t \geq 0, \forall z^t \in Z^t : \int_0^1 \int_{\theta^t} \int_{\pi^t} c_{it}(\pi^t, \theta^t, z^t) p_s^t(\pi^t, \theta^t) d\pi^t d\theta^t di = C_t(z^t),$$

and $n_{it}(\pi^t, \theta^t, z^t) \in \{0, \overline{n}\}$ for all $i, t, \pi^t, \theta^t, z^t$.[6] Without loss of generality, we can restrict the allocation to depend only on the individual's history for π and θ and not on the history of π and θ for all agents.

In order to make this problem convex, we now allow agents to randomize their labor supply: $\alpha_{it}\left(\pi^t, \theta^t, z^t\right)$ denotes the probability to work for agent i after history (π^t, θ^t, z^t). However, labor supply will turn out to be almost always deterministic. Hence these lotteries are only an analytical device and their presence does not affect the equilibrium. The planner problem is now to choose a sequence of functions $\left\{c_{it}(\pi^t, \theta^t, z^t), \alpha_{it}\left(\pi^t, \theta^t, z^t\right)\right\}$ to maximize:

$$\sum_{t \geq 0} \sum_{z^t \in Z^t} \beta^t p_z^t(z^t) \int_0^1 \int_{\theta^t} \int_{\pi^t} \mu_i \left[u(c_{it}(\pi^t, \theta^t, z^t)) - \theta_t \alpha_{it}(\pi^t, \theta^t, z^t) v(\overline{n})\right] p_s^t(\pi^t, \theta^t) d\pi^t d\theta^t di,$$

[6] Note that π and θ are individual-specific, but we do not index them by i because we integrate over them.

subject to:

$$\forall t \geq 0, \forall z^t \in Z^t : \qquad \overline{n} \int_0^1 \int_{\theta^t} \int_{\pi^t} \pi_t \alpha_{it}(\pi^t, \theta^t, z^t) p_s^t(\pi^t, \theta^t) d\pi^t d\theta^t di = N_t(z^t),$$
$$\forall t \geq 0, \forall z^t \in Z^t : \qquad \int_0^1 \int_{\theta^t} \int_{\pi^t} c_{it}(\pi^t, \theta^t, z^t) p_s^t(\pi^t, \theta^t) d\pi^t d\theta^t di = C_t(z^t),$$

and the requirement that $0 \leq \alpha_{it}(\pi^t, \theta^t, z^t) \leq 1$ for all $i, t, \pi^t, \theta^t, z^t$. We can separate this problem into two subproblems: first maximize the weighted sum of utilities of consumption, given the aggregate consumption endowment process:

$$\overline{U}(\{C_t\}) = \max_{\{c_{it}(\pi^t, \theta^t, z^t)\}} \sum_{t \geq 0} \sum_{z^t \in Z^t} \beta^t p_z^t(z^t) \int_0^1 \int_{\theta^t} \int_{\pi^t} \mu_i u(c_{it}(\pi^t, \theta^t, z^t)) p_s^t(\pi^t, \theta^t) d\pi^t d\theta^t di,$$

subject to:

$$\forall t \geq 0, \forall z^t \in Z^t : \int_0^1 \int_{\theta^t} \int_{\pi^t} c_{it}(\pi^t, \theta^t, z^t) p_s^t(\pi^t, \theta^t) d\pi^t d\theta^t di \leq C_t(z^t).$$

This is a standard risk-sharing problem with complete markets (as in Townsend, 1994). (This subproblem is relevant for us only in that it allows to have a representative agent.) Denote by $\lambda_t(z^t)$ the multiplier on the consumption constraint. Taking first order conditions with respect to $c_{it}(\pi^t, \theta^t, z^t)$ yields for all agents i, all time t and aggregate state z^t, and individual histories π^t, θ^t

$$\beta^t p_z^t(z^t) \mu_i u'(c_{it}(\pi^t, \theta^t, z^t)) p_s^t(\pi^t, \theta^t) = \lambda_t(z^t) p_s^t(\pi^t, \theta^t),$$

hence we obtain the standard risk-sharing rule: the MRS across dates and states is equalized between agents:

$$\frac{u'(c_{it}(\pi^t, \theta^t, z^t))}{u'(c_{ik}(\pi^k, \theta^k, z^k))} = \frac{\beta^k p_z^k(z^k)}{\beta^t p_z^t(z^t)} \frac{\lambda_t(z^t)}{\lambda_k(z^k)} = \frac{u'(c_{jt}(\widetilde{\pi}^t, \widetilde{\theta}^t, z^t))}{u'(c_{jk}(\widetilde{\pi}^k, \widetilde{\theta}^k, z^k))}.$$

Hence the solution satisfies $c_{it}(\pi^t, \theta^t, z^t) = f_i(C_t(z^t))$, where f_i is a function which depends only on i.

The second subproblem is the one which interests us in this paper. The planner minimizes the cost of providing the labor process $\{N_t\}$:

$$\overline{V}(\{N_t\}) = \min_{\{\alpha_{it}(\pi^t, \theta^t, z^t)\}} \sum_{t \geq 0} \sum_{z^t \in Z^t} \beta^t p_z^t(z^t) \int_0^1 \int_{\theta^t} \int_{\pi^t} \mu_i \theta_t \alpha_{it}(\pi^t, \theta^t, z^t) v(\overline{n}) p_s^t(\pi^t, \theta^t) d\pi^t d\theta^t di,$$

subject to:

$$\forall t \geq 0, \forall z^t \in Z^t : \overline{n} \int_0^1 \int_{\theta^t} \int_{\pi^t} \pi_t \alpha_{it}(\pi^t, \theta^t, z^t) p_s^t(\pi^t, \theta^t) d\pi^t d\theta^t di \geq N_t(z^t).$$

Let $\zeta_t(z^t)$ be the multiplier on the labor constraint. Taking the first-order condition with respect to $\alpha_{it}(\pi^t, \theta^t, z^t)$ yields:

$$\beta^t p_z^t(z^t)\mu_i\theta_t v(\overline{n}) - \zeta_t(z^t)\overline{n}\pi_t \begin{cases} < 0 \text{ if } \alpha_{it}(\pi^t, \theta^t, z^t) = 1, & \text{i.e. } n_{it}\left(\pi^t, \theta^t, z^t\right) = \overline{n}, \\ = 0 \text{ if } \alpha_{it}(\pi^t, \theta^t, z^t) \in (0,1), & \\ > 0 \text{ if } \alpha_{it}(\pi^t, \theta^t, z^t) = 0, & \text{i.e. } n_{it}\left(\pi^t, \theta^t, z^t\right) = 0, \end{cases} \tag{2.1}$$

The first-order condition for lottery choice implies that, if the distributions for π, θ, and μ are atomless, actual randomization (i.e. $\alpha_{it}(\pi^t, \theta^t, z^t) \in (0,1)$) is a zero-probability event. This is because:

$$\alpha_{it}(\pi^t, \theta^t, z^t) \in (0,1) \iff \frac{\zeta_t(z^t)}{\beta^t p_z^t(z^t)} \frac{\overline{n}}{v(\overline{n})} = \frac{\theta_{it}\mu_i}{\pi_{it}},$$

is a zero-probability event if the distributions H and p_s^t are atomless, since the left-hand side does not depend on i. The intuition for this result is that any lottery – any extraneous randomization device – can be replaced by actual exogenous randomness: the purification argument relies on the economic environment containing "as much risk" as any artificial lottery.[7] Intuitively, instead of rolling a dice and making his labor supply contingent on the outcome, a worker can decide beforehand that he will work when his random (relative) cost of working θ/π: by picking the threshold for θ/π, he obtains the same result as a lottery, and because of complete markets he can insulate his consumption from his labor supply.[8]

Define the aggregate wage rate w_t as the marginal utility of one unit of leisure over the marginal utility of one unit of consumption: $w_t(z^t) = \zeta_t(z^t)/(\beta^t p_z^t(z^t)\lambda_t(z^t))$.

[7] "Purification" refers to the fact that mixed strategies are actually degenerate, and hence equivalent to pure strategies with probability one in this environment. But even if the economy has no uncertainty, lotteries are not needed: workers can average over time, switching in and out of the labor force to choose the average time that they work (under continous time and continuous processes). This is however different than purification because the *dates* where an agent works may be indeterminate, even though the total time the agent works is not.

[8] Chang and Kim (2006a and 2006b) study quantitatively the extent to which this result is weakened when markets are incomplete.

Result 1: Labor Force Participation

From (2.1), we obtain the participation equation: agent i works in state (π^t, θ^t, z^t), i.e. $n_{it}\left(\pi^t, \theta^t, z^t\right) = \overline{n}$, if and only if the benefit is greater than the cost:

$$\lambda_t(z^t)w_t(z^t)\overline{n}\pi_{it} \geq \mu_i\theta_{it}v(\overline{n}). \tag{2.2}$$

This is the equation that we will estimate in Section 4. A more intuitive way to state this result is the following. Define the individual wage w_{it} as $w_{it} = w_t\left(z^t\right)\pi_{it}\overline{n}$, and the reservation wage w_{it}^R as $w_{it}^R = \frac{\theta_{it}\mu_i v(\overline{n})}{\lambda_t(z^t)}$. The participation equation (2.2) then amounts to $w_{it} \geq w_{it}^R$. The reservation wage w_{it}^R takes into account the idiosyncratic disutility of labor, but also the marginal utility of consumption, as measured by $\lambda_t\left(z^t\right)$, and the wealth of an agent, as measured by his Pareto-weight μ_i.

C. Aggregate Implications

Define aggregate employment $\widetilde{N}_t\left(z^t\right)$ as the physical quantity of aggregate hours worked, or equivalently as the number of agents agents employed times the workweek $\overline{n}$ (since there is no intensive margin):

$$\widetilde{N}_t\left(z^t\right) = \overline{n}\int_0^1\int_{\theta^t}\int_{\pi^t}\mathbf{1}_{\left[n_{it}\left(\pi^t,\theta^t,z^t\right)=\overline{n}\right]}p_s^t\left(\pi^t,\theta^t\right)d\pi^t d\theta^t di,$$

where $\mathbf{1}_A$ is a characteristic function. We can abstract from lotteries in this expression since the set of people who use lotteries is measure zero, and changing a function inside an integral on a zero-measure set does not change the value of the integral. Note that this is different from aggregate labor $N_t\left(z^t\right)$, which is defined in efficiency units. We focus on physical employment $\widetilde{N}_t$ because it is easier to measure empirically and it has been the focus of the literature, but we show below that our results can be easily reformulated to apply to efficiency-weighted hours N_t.

Rewrite the participation equation (2.2) as[9]

$$n_{it}\left(\pi^t, \theta^t, z^t\right) = \overline{n} \text{ if only if } x_{it} \overset{def}{=} \frac{\theta_{it}\mu_i}{\pi_{it}} \leq \frac{\overline{n}}{v(\overline{n})}\lambda_t\left(z^t\right)w_t\left(z^t\right) \overset{def}{=} x_t^*\left(z^t\right),$$

[9] It is a slight abuse to write "if and only if" since in the case of an equality, randomization occurs, and $n_t\left(\pi^t, \theta^t, z^t\right) = 0$ could happen.

where the variable x_{it} summarizes individual heterogeneity: we collapse various dimensions of heterogeneity – tastes, abilities, wealth – into a single variable x. Then the participation decision rule is summarized by a simple cutoff x_t^*, such that agent i works in period t if and only if $x_{it} \leq x_t^*$. To compute aggregate hours, define $G(\cdot)$ as the cumulative distribution function of $\log x_{it} = \log \frac{\theta_{it}\mu_i}{\pi_{it}}$. $G(\cdot)$ is constructed from the distribution H of Pareto-weights and from the invariant distribution of the Markov process for $s = (\pi, \theta)$, which we denote m:

$$\begin{aligned} G(x) &= \Pr\left[\log \frac{\theta\mu}{\pi} \leq x\right] \\ &= \Pr\left[\mu \leq \frac{\pi}{\theta} \exp x\right] \\ &= \int_\Theta \int_\Pi H\left(\frac{\pi}{\theta} \exp x\right) m(d\pi, d\theta). \end{aligned}$$

Then aggregate employment is:

$$\widetilde{N}_t\left(z^t\right) = G\left(\log x_t^*\left(z^t\right)\right) = G\left(\log\left(\frac{\overline{n}}{v(\overline{n})}\right) + \log \lambda_t\left(z^t\right) + \log w_t\left(z^t\right)\right). \tag{2.3}$$

Result 2: Aggregate Frisch Elasticity of Labor Supply

Equation (2.3) characterizes the Frisch labor supply schedule in this economy, which depends on the marginal utility of aggregate consumption $\lambda_t(z^t)$, and the marginal product of labor w_t. The Frisch elasticity of aggregate labor is:

$$\frac{\partial \log \widetilde{N}_t}{\partial \log w_t} = \frac{g\left(\log x_t^*\left(z^t\right)\right)}{G(\log x_t^*\left(z^t\right))}. \tag{2.4}$$

At the aggregate level, the Frisch elasticity of labor supply is a measure of homogeneity of the workforce at the margin (akin to a hazard rate). In this model, therefore, the elasticity can take any value, depending on the shape of the distribution of heterogeneity around the marginal worker. If the density g is high around $\log x_t^*(z^t)$, then there are many workers that are indifferent between working and not, and small fluctuations in the aggregate wage generate large aggregate fluctuations in labor supply.[10]

Relation to Rogerson (1988)

To see that Rogerson's (1988) infinite Frisch elasticity is a special case, consider this simple example. Assume that G is the uniform distribution over $[a - \varepsilon, a + \varepsilon]$ with density $\frac{1}{2\varepsilon}$. Any

[10] A similar formula obtains for the elasticity of efficiency-weighted hours. The only difference is that it incorporates the productivity of the marginal agent.

allocation where the fraction of the population that works is positive and less than one satisfies:

$$\left\{\log\left(\frac{\overline{n}}{v(\overline{n})}\right)+\log\lambda_t\left(z^t\right)+\log w_t\left(z^t\right)\right\}\in(a-\varepsilon,a+\varepsilon),$$

so that the Frisch Elasticity is:

$$\frac{1}{2\varepsilon}\left(\frac{\log\left(\frac{\overline{n}}{v(\overline{n})}\right)+\log\lambda_t\left(z^t\right)+\log w_t\left(z^t\right)-(a-\varepsilon)}{2\varepsilon}\right)^{-1}>\frac{1}{2\varepsilon},$$

which obviously becomes infinite as ε tends to zero, that is as the economy gets homogeneous. In this case, all workers are marginal, hence aggregate labor supply is infinitely elastic.

Equivalence with a representative agent economy

Our economy has the same aggregate implications as a representative agent economy with divisible labor as in Lucas and Rapping (1969), with aggregate utility function

$$\begin{aligned} W\left(\{C_t,N_t\}\right) &= \overline{U}\left(\{C_t\}\right)-\overline{V}\left(\{N_t\}\right)\\ &= \sum_{t\geq 0}\sum_{z^t\in Z^t}\beta^t p_z^t(z^t)\left(U(C_t(z^t))-V(N_t(z^t))\right), \end{aligned}$$

where $U(\cdot)$ is defined as:

$$\begin{aligned} U(x) &= \max_{\{c_i(\pi,\theta)\}}\int_0^1\int_\theta\int_\pi \mu_i u(c_i(\pi,\theta))m(d\pi,d\theta)di,\\ s.t. &: \int_0^1\int_\theta\int_\pi c_i(\pi,\theta)m(d\pi,d\theta)di\leq x, \end{aligned}$$

and $V(\cdot)$ is defined as:

$$\begin{aligned} V(y) &= \min_{\{\alpha_i(\pi,\theta)\}} v(\overline{n})\int_0^1\int_\theta\int_\pi \mu_i\theta\alpha_i(\pi,\theta)m(d\pi,d\theta)di,\\ s.t. &: \overline{n}\int_0^1\int_\theta\int_\pi \pi\alpha_i(\pi,\theta)m(d\pi,d\theta)di\geq y. \end{aligned}$$

This equivalence was also noted by Mulligan (2001). Given the optimal decision rule (i.e. $\alpha_{it}=1$ iff $\log\theta_{it}\mu_i/\pi_{it}\leq\log x_t^*$), this can be shortened as

$$V(y)=v(\overline{n})E_{\mu,\theta,\pi}\left[\mu\theta\times 1_{\log\frac{\mu\theta}{\pi}\leq\log x^*(y)}\right],$$

where the expectation is taken under the ergodic measure, and $x^*(y)$ is implicitely defined by:

$$\overline{n}E_{\mu,\theta,\pi}\left[\pi\times 1_{\log\frac{\mu\theta}{\pi}\leq\log x^*(y)}\right]=y.$$

Along the optimal labor supply allocation, there is a one-to-one mapping between physical hours and efficiency-weighted labor; in the notation above y is efficiency-weighted and physical hours is:

$$\widetilde{N}(y) = \overline{n} E_{\mu,\theta,\pi} \left[1_{\log \frac{\mu\theta}{\pi} \leq \log x^*(y)} \right].$$

Extension to the case of fixed effects in θ and π

We have assumed that $s_{it} = (\theta_{it}, \pi_{it})$ follows a Markov process. This rules out the case of fixed effects, which may be empirically relevant. However, it is straightforward to incorporate fixed effects in our analysis, both for θ_{it} and π_{it}: write $\theta_{it} = \theta_i \varepsilon_{it}^{\theta}$ and $\pi_{it} = \pi_i \varepsilon_{it}^{\pi}$ where $(\varepsilon_{it}^{\pi}, \varepsilon_{it}^{\theta})$ follows a Markov process. Since we already have fixed effects through the Pareto weight μ_i, the results 1 and 2 also hold exactly in this case.

The importance of distributional assumptions

In this section, we show the important role of the distributions used to calibrate tastes, abilities and wealth. To make this comparison as transparent as possible, we assume tastes are homogeneous, wealth is equal across households, and we choose a stationary distribution F for the logarithm of individual productivity $\log \pi_i$. We consider several distributions, each of which is characterized by two parameters. We choose these two parameters to obtain $E(\log \pi) = 0$ and $Var(\log \pi) = 0.6964$. These two moments are chosen to replicate the observed properties of the residuals in the wage regression, i.e. the idiosyncratic wage shocks. (The precise number for the variance is taken from Chang and Kim 2006b.) We then choose the cutoff π^* to match an employment rate of 60% on average. Table 1 reports the implied elasticities for various distributions.

Distribution of $\log \pi$	Implied Aggregate Frisch Elasticity
Normal	0.77
Logistic	0.87
Uniform	0.40
Pareto (distribution of π)	1.20
Mixture of two normals ($r = 1$)	0.98
Mixture of two normals ($r = 10$)	7.03

Table 1: Model-implied Frisch Elasticity in Complete Markets, for various distributions. All the distributions have the same variance and have mean 0, with an employment rate of 60%.

The last two rows of the table refer to the case where the distribution is a mixture of two normals, i.e.:

$$F(x) = \alpha\Phi\left(\frac{x-\mu_1}{\sigma_1}\right) + (1-\alpha)\Phi\left(\frac{x-\mu_2}{\sigma_2}\right),$$

where Φ is the standard normal cumulative distribution function. These two examples share the same parameters, except for the ratio of standard deviations $r = \sigma_2/\sigma_1$.[11]

We find it interesting that reasonable variations in the precise shape of individual heterogeneity, as represented by these distribution functions, can change the implied elasticity by a large amount, even when the variance is the same. Even without considering mixtures, the elasticity ranges from 0.40 for the uniform distribution to 1.20 for the Pareto distribution. With mixtures, any elasticity can be generated.[12] From a mathematical point of view, this should not be surprising: the elasticity is determined by a specific value of the hazard rate, which has no direct relation to the three conditions used for the calibration. These computations suggest that the shape of heterogeneity is quantitatively important. In section 3, we exploit the tractability of complete markets to develop an empirical model that allows for richer heterogeneity.

The importance of microdata

In this model, a representative agent exists and the usual first-order condition which equates the marginal utility of leisure over the marginal utility of consumption to the wage holds,

[11] Clearly, there are more parameters than moments. We set arbitrarily $\alpha = .50$ and $\mu_1 = -0.8$, and choose μ_2 to satisfy $E(\log \pi) = 0$. Next, for a given r, we choose σ_1 (and thus $\sigma_2 = r\sigma_1$) to match $Var(\log \pi)$.

[12] One might view mixtures as unrealistic. However, some dimensions of heterogeneity are discrete (e.g. the number of children), which naturally generates mixtures.

for some utility function which reflects the underlying heterogeneity: $V'(N_t(z^t))/U'(C_t(z^t)) = w_t(z^t)$ for all t, z^t. Mulligan (2001, 2002) concludes that there is little point in estimating the aggregate labor supply elasticity with microeconomic data, and he uses the aggregate marginal condition to estimate the aggregate labor supply elasticity.

While the equivalence makes his approach appealing, we think it has also some drawbacks. First, it is necessary to specify a functional form for the marginal utility of consumption, and to measure the marginal product of labor. The former has proved problematic in the asset pricing literature, and the later is hard because of the compositional effect in average wage or productivity series (Solon, Barsky and Parker 1994). Second, our procedure is robust to labor supply shocks which many authors deem important (e.g., Chari, Kehoe and McGrattan 2007): by using microeconomic data we are able to take into account these aggregate shocks in our estimation using time effects. Third, our procedure allows us to 'test' whether the aggregation story fits with the micro data: we can measure how many workers are 'marginal' and how much they contribute to aggregate fluctuations. Finally, we are able to evaluate the elasticity at any point in time. This allows us to check if the elasticity varies over time. The model suggests this may be true: after a long expansion, many potential workers enter the workforce, and the reservation wage of the remaining potential workers may be high. This could make the elasticity of labor supply become lower after a long boom such as the late 1990s, and conversely we might expect the elasticity of labor supply to become high in recessions.[13] (Of course, since this depends on the shape of the distribution, which is unrestricted by the theory, the effect could in principle go either way.)

D. Cross-Sectional Implications and Marginal Workers

Since the participation decision (2.2) is almost never interior, only the marginal worker (who is just indifferent between working and not working) would react to a marginal change in aggregate employment (or the aggregate wage). Hence ex post – after all idiosyncratic shocks

[13] For instance, the New York Times reported on December 20, 1999, that "As labor pools shrinks, a new supply is tapped". The article discusses how some individuals who were out of the labor force (students and retired people) were agressively recruited by expanding businesses. On the other hand, then-Chairman of the Federal Reserve Alan Greenspan worried at the same time about the size of the "shrinking pool of available workers" (Remarks on April 5, 2000).

are realized – the marginal effect of w_t is nil except for this marginal worker. Hence only these workers are sensitive to aggregate shocks. However, this ex-post prediction has no content unless we "identify" the marginal workers.

A natural solution is to consider the marginal effect of w_t, *before* x_{it} *is known.* That is, consider our one-dimensional measure of idiosyncratic risk $x_{it} \equiv \theta_{it}\mu_i/\pi_{it}$, and suppose we observe X_{it}, a variable which is correlated with x_{it}. Our model immediately yields the following formula for the marginal effect of a change in aggregate wage (i.e. the elasticity of the probability of working to the wage rate), as a function of X_{it} :

Result 3: Heterogeneous Elasticities

$$\frac{\partial \log \Pr\left[n_{it}=1 \mid X_{it}\right]}{\partial \log w_t} = \frac{g\left(\log \lambda_t\left(z^t\right)+\log w_t\left(z^t\right) \mid X_{it}\right)}{G\left(\log \lambda_t\left(z^t\right)+\log w_t\left(z^t\right) \mid X_{it}\right)} = \frac{g\left(\log x_t^*|X_{it}\right)}{G\left(\log x_t^*|X_{it}\right)},$$

where $g(x|y)$ (resp. $G(x|y)$) is the conditional probability density function (resp. cumulative distribution function) of x_{it} given $X_{it}=y$.

This equation implies that the *level* change in the probability is

$$\frac{\partial \Pr\left[n_{it}=1 \mid X_{it}\right]}{\partial \log w_t} = g\left(\log x_t^*|X_{it}\right).$$

Assuming that $G(x|y)$ is increasing in y, (i.e. that "x_{it} is monotone in X_{it}"), and assuming that the density g, which is defined on all $\mathbb{R}$, is single-peaked, we can see that this change in probability $\frac{\partial \Pr[n_{it}=1|X_{it}]}{\partial \log w_t}$ will be low for very high or very low X_{it}, for which agents are almost surely working or not working, and high for intermediate values of X_{it}. The marginal effect $\frac{\partial \log \Pr[n_{it}=1|X_{it}]}{\partial \log w_t}$ will also be low for high X_{it}, but it may be high for low X_{it} because the denominator $G\left(\log x_t^*|X_{it}\right)$ is small. This is illustrated in Figure 1.

Importantly, the aggregate elasticity, given in (2.4), is a weighted average of these marginal effects. Denote by F the cumulative distribution function of X_{it}, then since $\widetilde{N}_t = \int \Pr(n_{it} = 1|X_{it})dF(X_{it})$, we have

$$\begin{aligned}
\frac{\partial \log \widetilde{N}_t}{\partial \log w_t} &= \int \frac{\partial \log \Pr\left[n_{it}=1 \mid X_{it}\right]}{\partial \log w_t} \frac{\Pr\left[n_{it}=1 \mid X_{it}\right]}{\widetilde{N}_t} dF\left(X_{it}\right), \\
&= \frac{1}{\widetilde{N}_t} \int \frac{\partial \Pr\left[n_{it}=1 \mid X_{it}\right]}{\partial \log w_t} dF\left(X_{it}\right).
\end{aligned}$$

The first line states that the aggregate elasticity is a weighted sum (by the probability to working) of individual marginal effects. The second line states that the elasticity is also the

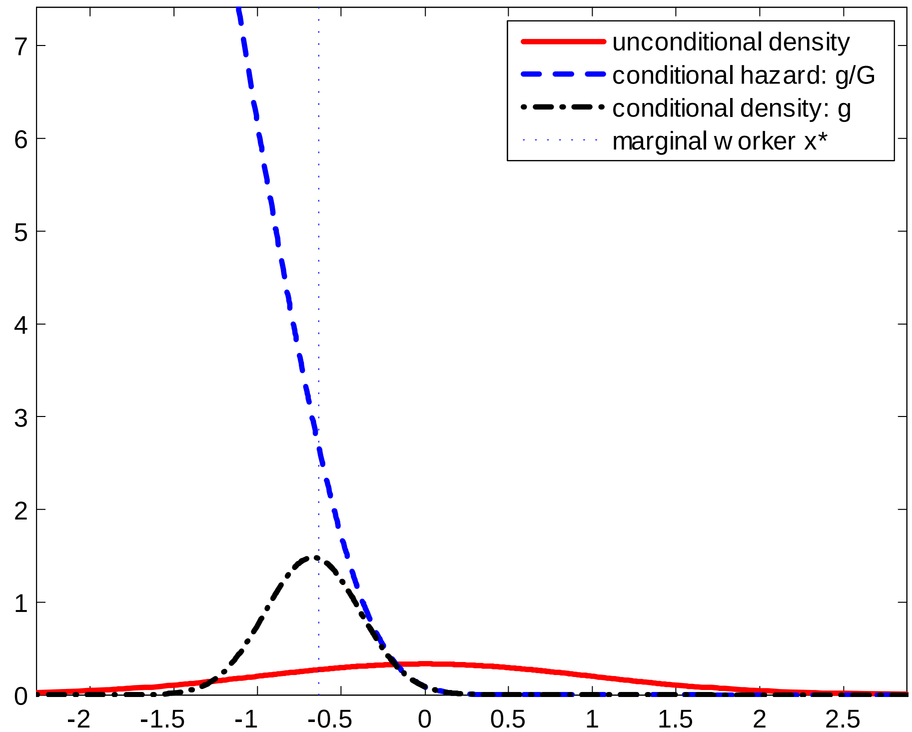

Figure 1: **The dashed line represents the marginal effects** $\partial \log \Pr[n_{it} = 1 \mid \pi_{it-1}] / \partial \log w_t$. **The dashed-dotted line represents the level effects** $\partial \Pr[n_{it} = 1 \mid \pi_{it-1}] / \partial \log w_t$. **The solid line represents the unconditional density of** $\log \pi_{it-1}$.

sum of the level effect on probabilities. Hence if there are enough "marginal workers", for whom the elasticity is high, the aggregate elasticity will be high. Clearly, the cross-sectional prediction is closely related to our aggregate implications, as the similarity of the formulas suggests.

In practice, we will use estimated fixed effects or covariates (i.e. observable characteristics) as X_{it}. However, X_{it} could also include lagged employment n_{it}, or lagged values of x_{it}, as in the example below.

Are the marginal workers constant?

An extreme version of our theory is that marginal workers are constant - they are a fixed set of people. This is the case when π and θ are fixed over time. The workers whose π and θ make them close to the reservation wage are marginal. In this case, the people with high π or low θ always work, the people with low π or high θ never work, and intermediate π or θ people are the only ones whose employment changes over time. This version of the theory makes sharp predictions on who adjusts over the business cycle.

The other extreme version is that π and θ are i.i.d. and agents have equal wealth (i.e. the same μ_i). In this case, all workers are equally likely to enter or exit the workforce, and the theory makes no prediction regarding who adjusts over the business cycle.

Importantly, whether marginal workers are the same people or not over time is irrelevant for our aggregate estimate. To put it another way, what matters to our "marginal homogeneity" argument is the *total* (unconditional) amount of heterogeneity at the margin, which can come from either fixed effects, covariates, or idiosyncratic shocks. (In particular, while it is reasonable to expect that idiosyncratic shocks are approximately log-normal, there is no reason to expect this for fixed effects or covariates.) The downside is that unless we are specific about the stochastic structure of idiosyncratic shocks, it is hard to generate predictions regarding the pattern of marginal effects and individual sensitivities to aggregate shocks.

As a more realistic example, consider the case where the only heterogeneity is in productivity π_{it}. Productivity follows an AR(1) process: $\log \pi_{it} = \rho \log \pi_{it-1} + \sigma \varepsilon_t$. We use $X_{it} = \pi_{it-1}$ as a predictor of π_{it}. We compute the marginal effect of a change in w_t on the probability of working at t given π_{it-1}. Figure 1 shows the heterogeneity in marginal effects $\frac{\partial \log \Pr[n_{it}=1|X_{it}]}{\partial \log w_t}$: the dotted line shows that most agents have small marginal effects, while a few agents with the lowest wages have high marginal effects. However, the level changes in probabilities $\frac{\partial \Pr[n_{it}=1|X_{it}]}{\partial \log w_t}$ (i.e. the marginal effects weighted by their actual probability of employment) are hump-shaped, so that the marginal worker is actually in the middle of the distribution, where both the marginal effect and the employment probability are substantial. The aggregate elasticity is the integral of the dashed-dotted line, with the distribution given by the full line.

In a previous draft (Gourio and Noual 2006), we used PSID data to test the prediction that marginal people have a more cyclical labor supply. We found that people with low schooling, low current hours, low family income, or low wages are more procyclical, when one runs a regression of growth rate of hours of each group on the growth rate of aggregate hours. When one runs this equation in levels (i.e. change in average hours on the change in aggregate average hours), the results yield in general a hump-shape. This is qualitatively consistent with the results above. (See Appendix D for more details).

3 Estimation Method

We first discuss the equation that we estimate, then we discuss the econometric method, the data and exact specification, and finally the elasticities we measure. Section 4 presents the results.

A. The participation and the wage equations

We use a balanced panel of I individual agents over T periods. We observe employment $n_{it} \in \{0, 1\}$, and if $n_{it} = 1$ we also observe wages w_{it}. Our model asserts that $n_{it} = 1$ if and only if $w_{it} \geq w_{it}^R$: the reservation wage w_{it}^R is never observed, while the wage rate w_{it} is observed only for workers.

In our model, the reservation wage is $w_{it}^R = \theta_{it}\mu_i v\left(\overline{n}\right)/\lambda_t(z^t)$ and the wage rate is $w_{it} = w_t\pi_{it}\overline{n}$. To bring this model to the data we need to specify empirical processes for θ_{it} and π_{it}. We generalize our model slightly to take into account observable factors which affect wages on one side, and reservation wages on the other side. While this feature is not present in our theoretical model, it is easy to incorporate it, at a notational cost. To model the unobserved heterogeneity in θ_{it} and π_{it}, we assume that they follow (in log) a process with an individual fixed effect and an iid shock. Formally, productivity and taste satisfy the two assumptions:

$$(A1) \quad : \quad \log \pi_{it} = \widetilde{a}_i + \mathbf{x}_{it}\gamma + u_{it},$$
$$(A2) \quad : \quad \log \theta_{it} = \widetilde{a}_i^R + \mathbf{y}_{it}\eta + e_{it},$$

where $\mathbf{x}$ and $\mathbf{y}$ are vectors of observable characteristics. We assume that the vector (u_{it}, e_{it}) is independent across i and t and jointly normal. In section 5.A, we show that the i.i.d. assumption does not affect our results significantly. The participation equation is (see equation 2.2):

$$\begin{aligned} \log w_{it} &\geq \log w_{it}^R, \\ \text{i.e. } \log w_t + \log \pi_{it} + \log \overline{n} &\geq -\log \lambda_t + \log \theta_{it} + \log v\left(\overline{n}\right) + \log \mu_i, \\ b_t + a_i + \mathbf{x}_{it}\gamma + u_{it} &\geq b_t^R + a_i^R + \mathbf{y}_{it}\eta + e_{it}, \end{aligned} \tag{3.1}$$

with $a_i^R = \widetilde{a}_i^R + \log v(\overline{n}) + \log \mu_i$, $a_i = \log \overline{n} + \widetilde{a}_i$, and we have replaced the aggregate wage and the disutility of labor over marginal utility of wealth by time effects, which also absorb the constant,[14] with $b_t^R = -\log \lambda_t$. The wage equation reads:

$$\begin{aligned} \log w_{it} &= \log w_t + \log \pi_{it} + \log \overline{n}, \\ &= \log w_t + a_i + \mathbf{x}_{it}\gamma + u_{it} + \log \overline{n}, \\ &= b_t + a_i + \mathbf{x}_{it}\gamma + u_{it}. \end{aligned} \tag{3.2}$$

These specifications allow for unobservable nonparametric heterogeneity with fixed effects. In this sense, the heterogeneity is unrestricted. On the other hand the fact that (u_{it}, e_{it}) is i.i.d. and normal is restrictive. However, some persistence can be contained in the vectors of observable covariates $\mathbf{x}_{it}, \mathbf{y}_{it}$. It might be interesting to relax this assumption using simulation methods (for instance as in Hyslop 1999).

B. **Econometric method**

We assume a joint normal distribution for transitory shocks on wages and reservation wages (u_{it}, e_{it}):

$$\begin{pmatrix} u_{it} \\ e_{it} \end{pmatrix} \overset{i.i.d.}{\sim} N\left(0, \Sigma\right) \text{ with } \Sigma = \begin{pmatrix} \sigma_u^2 & \sigma_{ue} \\ \sigma_{ue} & \sigma_e^2 \end{pmatrix}.$$

The likelihood of our observations on $\{n_{it}, w_{it}\}$ is:

$$L = \prod_{i=1}^{N} \prod_{t=1}^{T} \left\{ \Pr\left[n_{it} = 0 \mid \mathbf{x}_{it}, \mathbf{y}_{it}\right] \right\}^{1-n_{it}} \left\{ \phi\left(w_{it}, n_{it} = 1 \mid \mathbf{x}_{it}, \mathbf{y}_{it}\right) \right\}^{n_{it}}, \tag{3.3}$$

where ϕ stands for the joint density of wages w_{it} and participation n_{it}, which can be recovered from the joint density of shocks u_{it} and e_{it}. Identification of the parameters $\sigma_u^2, \sigma_{ue}, \sigma_e^2, \gamma, \eta$, and

[14] The variable λ_t is not observed. The aggregate wage w_t may be observed, but there is a composition bias, so we also prefer to use time effects.

$\left\{a_i^R\right\}_{i=1}^N, \{b_t\}_{t=1}^T, \left\{b_t^R\right\}_{t=1}^T$ follows from the maximization of the likelihood. We experimented with this likelihood maximization, using the tools developed by Greene (2001), who shows how to exploit the sparsity of Hessians matrices to estimate a large number of fixed effects. However, the likelihood is not concave, which makes it hard to attain the global maximum. For this reason, we use instead the Heckman (1979) two-step estimator.[15] We still use the tools of Greene (2001) to perform the first step (the Probit model).

Rephrasing the first-order condition (2.2) in terms of our statistical model,

$$\begin{aligned}
\Pr\left[n_{it}=1\right] &= \Pr\left[w_{it}^R \leq w_{it}\right] = \Pr\left[b_t^R + a_i^R + \mathbf{y}_{it}\eta + e_{it} \leq b_t + a_i + \mathbf{x}_{it}\gamma + u_{it}\right], \\
&= \Pr\left[e_{it} - u_{it} \leq b_t + a_i + \mathbf{x}_{it}\gamma - \left(b_t^R + a_i^R + \mathbf{y}_{it}\eta\right)\right], \\
&= \Phi\left(\frac{b_t + a_i + \mathbf{x}_{it}\gamma - \left(b_t^R + a_i^R + \mathbf{y}_{it}\eta\right)}{\sigma}\right), \qquad (3.4)
\end{aligned}$$

where $\sigma^2 = Var\left(e_{it} - u_{it}\right) = \sigma_u^2 + \sigma_e^2 - 2\sigma_{ue}$, and Φ is the cumulative function of the standard normal distribution. This participation equation (3.4) is a Probit model which can be estimated by MLE. The Probit allows to recover estimates of $\left\{\left\{\frac{b_t - b_t^R}{\sigma}\right\}_t, \left\{\frac{a_i - a_i^R}{\sigma}\right\}_i, \frac{\gamma}{\sigma}, \frac{\eta}{\sigma}\right\}$.

However, to identify σ separately from the other parameters γ, η, and $\{a_i\}_{i=1}^N$, $\left\{a_i^R\right\}_{i=1}^N$, $\{b_t\}_{t=1}^T$, $\left\{b_t^R\right\}_{t=1}^T$, the wage equation (3.2) needs to be estimated as well. Identifying σ is crucial for our purpose since the slope of the aggregate labor supply curve depends on its value. Therefore, after the probit has been estimated, we estimate the wage regression for workers, controlling for selection. As a result, our second step consists of the following OLS regression:

$$\log w_{it} = b_t + a_i + x_{it}\gamma + \frac{\sigma_u^2 - \sigma_{ue}}{\sigma}\frac{\phi\left(c_{it}\right)}{\Phi\left(c_{it}\right)} + u_{it}, \qquad (3.5)$$

where the last term is the inverse Mills ratio, evaluated at the index c_{it} which determines participation:

$$c_{it} = \frac{b_t + a_i + \mathbf{x}_{it}\gamma - \left(b_t^R + a_i^R + \mathbf{y}_{it}\eta\right)}{\sigma}. \qquad (3.6)$$

In this second step, all parameters are identified from an exclusion restriction: some variables in $\mathbf{y}_{it}$ are not in $\mathbf{x}_{it}$, that is there are determinants of the reservation wages (such as the age of children) that have no effect on market wages: as a result, the Mills ratio is not colinear to the other determinants of wages.

[15] Matlab programs which perform our estimation are available upon request.

From our estimates of $\widehat{\frac{\gamma}{\sigma}}$ in the first step (probit), and our estimates of $\widehat{\gamma}$ in the second step (wage regression), we recover σ. Because $\mathbf{x}_{it}$ is not a scalar, we construct a minimum distance estimator which minimizes a weighted average of $\left(\widehat{\gamma}_k - \widehat{\sigma}\widehat{\frac{\gamma_k}{\sigma}}\right)^2$ for all variables k that we take as regressors for wages. From our estimate of σ, we then recover all other parameters.

C. Data and Specification

We use the National Longitudinal Survey of Youth 1979 (NLSY 79). (See Appendix A for a more detailed description of this well-known data set.) This gives us a panel with $N = 5571$ agents and $T = 168$ months. Our sample starts in January 1979 and ends in December 1992.[16] To estimate the probit with fixed effects, we need to exclude agents who are either always working or never working. However, we take these agents into account in our computations of aggregate elasticities.

We use the following variables in our specification. The vector $\mathbf{x}_{it}$ includes determinants of wages: experience, experience squared, and schooling. The vector $\mathbf{y}_{it}$ includes determinants of reservation wages: a dummy for the marital status (interacted with gender), a dummy if the youngest child is less than 2 years old, a dummy if the yougest child is between 3 and 6 years old, and a dummy if the youngest child is between 7 and 14 years old[17] and a dummy for the answer to a question: "do health problems limit the amount or type of work [you] can perform?". In unreported results, we tried to add more covariates, including interactions, but the aggregate results did not change markedly; however, this might require further investigation.

D. Measurement of Elasticities

This subsection explains in detail how we use our estimates to compute macroeconomic or group-level elasticities. This is important since our model provides the explicit mapping between micro parameters and the macro elasticity. Consider the latent index c_{it} that determines participation:

$$c_{it} \equiv \frac{b_t + a_i + \mathbf{x}_{it}\gamma - \left(b_t^R + a_i^R + \mathbf{y}_{it}\eta\right)}{\sigma}.$$

[16] Our data cover 1979-1998 in fact, yet we are much less confident about the wage data after 1993, due to a change in methodology in the NLSY survey.

[17] The children dummies are set to 1 only for the female respondents.

The index c_{it} is the difference between the predicted wage rate w_{it} and reservation wage rate w_{it}^R, scaled by the variance of i.i.d. shocks. It is estimated from the first step of our estimation, i.e. the probit model for participation. Predicted aggregate employment is: $\widehat{N}_t = \sum_{i=1}^I \Pr\left[n_{it}=1\right] = \sum_{i=1}^I \Phi\left(c_{it}\right).$

To compute the aggregate elasticity of labor supply, we evaluate the derivative with respect to b_t, since b_t captures the aggregate wage in our equation:

$$\frac{\partial \log \widehat{N}_t}{\partial \log b_t} = \frac{1}{\sigma} \frac{\sum_{i=1}^I \phi\left(c_{it}\right)}{\sum_{i=1}^I \Phi\left(c_{it}\right)}. \tag{3.7}$$

Importantly, this definition of the aggregate elasticity does not always seem to be the measure reported by researchers, who sometimes report the response of the average or median individual, i.e. $\frac{1}{\sigma}\phi(\bar{c})/\Phi\left(\bar{c}\right)$ where $\bar{c}$ is the average of the c_{it}. This difference may be important, as we demonstrate in Section 4. Note how

$$\frac{1}{\sigma} \frac{\sum_{i=1}^I \phi\left(c_{it}\right)}{\sum_{i=1}^I \Phi\left(c_{it}\right)} = \frac{1}{\sigma} \frac{\sum_{i=1}^I \frac{\phi(c_{it})}{\Phi(c_{it})}\Phi(c_{it})}{\sum_{i=1}^I \Phi\left(c_{it}\right)}$$

is simply a weighted sum of hazard rates (or marginal effects, once scaled by σ^{-1}). One can interpret this formula as an average of marginal workers: for each value of the index c_{it}, there is a continuum of workers and $\sigma^{-1}\phi\left(c_{it}\right)$ of them are "marginal". The elasticity is the weighted sum of these numbers of marginal workers. When i.i.d. shocks are larger, as measured by σ, there is more heterogeneity and the aggregate elasticity falls.[18]

We also present results by groups: in this case, the index i in each of the two sums in (3.7) ranges over all the i in the group rather than $i = 1$ to I. For instance, the elasticity of men is:

$$\frac{\partial \log \widehat{N}_t^{men}}{\partial \log b_t} = \frac{1}{\sigma} \frac{\sum_{i \in men} \phi\left(c_{it}\right)}{\sum_{i \in men} \Phi\left(c_{it}\right)}.$$

Obviously, we can then break down the aggregate elasticity into the elasticities of the different groups, weighted by their shares in employment. For instance, if the decomposition is by gender, we have:

$$\frac{\partial \log \widehat{N}_t}{\partial \log b_t} = \frac{\partial \log \widehat{N}_t^{men}}{\partial \log b_t} \frac{\widehat{N}_t^{men}}{\widehat{N}_t} + \frac{\partial \log \widehat{N}_t^{women}}{\partial \log b_t} \frac{\widehat{N}_t^{women}}{\widehat{N}_t}.$$

[18] This result is not general, but depends on the normal distribution assumption.

4 Empirical Results

This section presents the empirical results obtained from the estimation procedure on our NLSY sample. We first present our estimated wage and participation equations. We next compute the Frisch elasticity implied by the model and we illustrate how the elasticity differs across groups formed on observable characteristics such as gender or schooling. We also discuss how this Frisch elasticity varies over the business cycle. Finally, we describe the "marginal workers" and their importance for aggregate fluctuations.

A. Estimation results

Table 2 reports the coefficients of the wage regression (3.5). (All coefficients and their standard errors are expressed in percentages to facilitate the interpretation.) The return to one year of schooling is around 10%, consistent with the usual results of Mincer regressions. The coefficients on experience and experience squared are also similar to standard results about the determinants of earnings (e.g. Heckman, Lochner and Todd 2003).

	point estimate (×100)	standard error (×100)
experience	12.16	0.18
experience squared	-0.26	0.01
schooling	9.88	0.09

Table 2: Estimates of coefficients γ
on observables $\mathbf{x}_{it}$ in the wage regression for w_{it}

Table 3 reports our estimates of the determinants of reservation wages. The results are sensible: being married raises the reservation wage for a woman, but it decreases the reservation wage for men. Having a medical condition that limits the type or amount of work one can perform raises the reservation wage by 6.3%, a significant amount. Finally, the reservation wage is much higher for women with young children, and the younger the children the higher the reservation wage.

	point estimate (×100)	standard error (×100)
married (men)	-4.63	0.13
married (women)	3.31	0.08
health limit	6.33	0.17
woman with youngest kid 0-2 years	17.61	0.46
woman with youngest kid 3-6 years	8.39	0.24
woman with youngest kid 7-14 years	3.86	0.16

Table 3: Estimates of coefficients η

on observables $\mathbf{y}_{it}$ in the reservation wage w_{it}^R

An important determinant of wages and reservation wages is unobserved permanent heterogeneity: this is captured by the fixed effects a_i, a_i^R (for wages and reservation wages respectively). Since labor force participation depends on the net effect $a_i - a_i^R$, we present its histogram in figure 2. There is a substantial amount of permanent heterogeneity that cannot be attributed to observables: the standard deviation of $\left(a_i - a_i^R\right)$ is 12.65%. Figure 2 also shows that this distribution is more concentrated than a normal distribution of same mean and variance (represented as 'fitted normal'). Note that we can recover from these estimates the distribution of π or of θ and μ; but we can not identify separately the fixed effect in θ_{it} and μ_i.

Finally, a crucial parameter is σ, the standard deviation of transitory shocks to wages and reservation wages w_{it}^R and w_{it}. (Recall $\sigma = Std\left(e_{it} - u_{it}\right)$.) We estimate $\sigma = 17.63\%$, with a standard error of 0.26%. Hence, i.i.d. shocks are large: their standard deviation is 50% larger than the standard deviation of fixed effects. We also report in Table 4 the structure of shocks to wages and reservations wages.

	point estimate (×100)
σ_u^2	12.40
σ_e^2	15.31
σ_{ue}	12.30

Table 4: Estimates of the variance of shocks

$\sigma^2 = Var\left(e_{it} - u_{it}\right) = \sigma_u^2 + \sigma_e^2 - 2\sigma_{ue}$

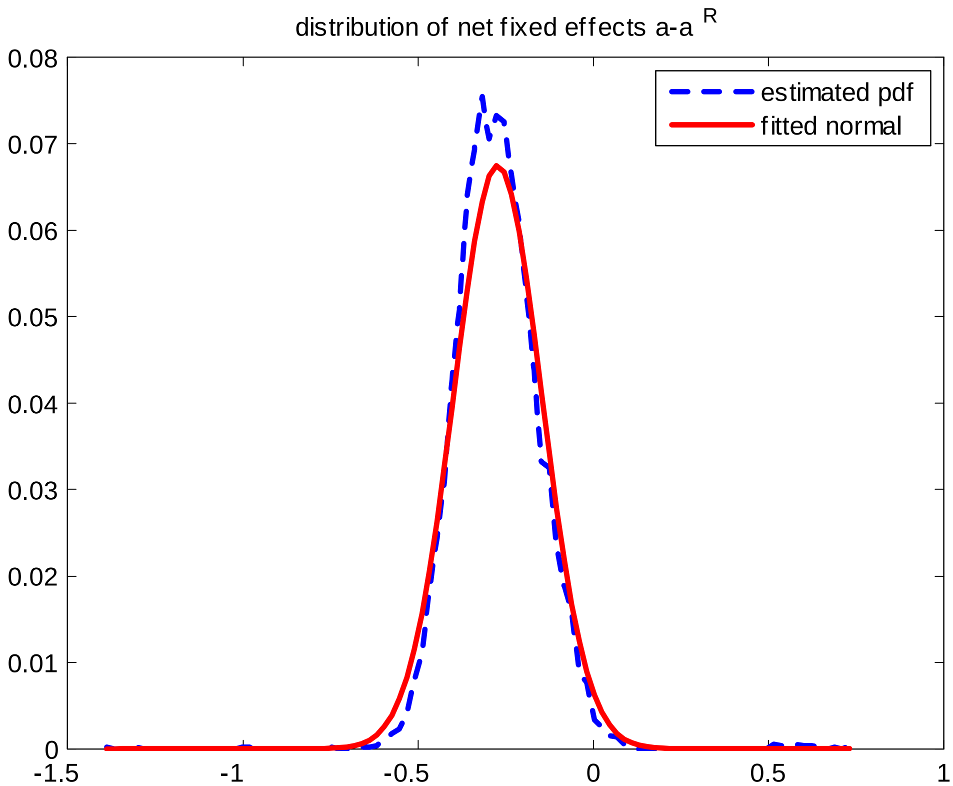

Figure 2: **The dashed line is the distribution of net fixed effect** $a_i - a_i^R$: a_i is a fixed effect in the individual wage w_{it}, and a_i^R is a fixed effect in the reservation wage w_{it}^R. The full line is a normal distribution with the same mean and variance as the dashed line.

In terms of goodness of fit, the average probability of correct prediction of employment status n_{it} is 77%.[19] Unsurprisingly given that we include time effects, the model fits aggregate employment very well. Overall, our results in these dimensions are consistent with the existing literature. We now turn to the aggregate implications of these estimates, using the measures that are implied by our theoretical analysis.

B. **The elasticity of labor supply**

Figure 3 depicts the Frisch aggregate elasticity at each date, together with the plus and minus two-standard errors bands. Figure 4 plots the aggregate Frisch elasticity together with aggregate employment. The employment series has two noticeable features. First, it has an important seasonal component, especially in the early part of the sample. Second, it trends up. These features are due to the changing age composition of our data. Because the NLSY follows some initial cohorts over time (people aged between 14 and 22 in 1979), their labor supply grows as they age. The Frisch elasticity is defined using our formula above:

$$\frac{\partial \log \widehat{N}_t}{\partial \log b_t} = \frac{1}{\sigma} \frac{\sum_{i=1}^{I} \phi\left(c_{it}\right)}{\sum_{i=1}^{I} \Phi\left(c_{it}\right)}, \tag{4.1}$$

where c_{it} is our estimate for the latent index that determines participation (i.e. $n_{it} = \overline{n}$ if $c_{it} \geq \varepsilon_{it}$ where $\varepsilon_{it} \sim N(0,1)$, see Section 3D). This elasticity is rather precisely estimated. The elasticity at the median date (July 1985) is 1.27, and the average elasticity over the whole sample is 1.50, which is higher than estimates based on the intensive margin, but still lower than the number of 3 or 4 required for macroeconomics (King and Rebelo 1999, Prescott 2006). Note the crucial role of σ in the expression (4.1); σ determines the amount of iid heterogeneity: the lower the heterogeneity, the higher the elasticity, and the Rogerson (1988) limit is obtained as $\sigma \to 0$. If there is measurement error in the wage, which seems likely, our estimate of the Frisch elasticity is thus likely downward biased. As we noted already, our measure of the elasticity differs from the marginal effects for the median (or average) agent which is often reported in the discrete choice literature. For instance, the marginal effect of w_t at the median date in our

[19] This measure has been suggested by Ben-Akiva and Lerman (1985). Another measure of goodness of fit, Efron's (1978), is 38% in our case. Efron's measure is close to an R^2, since it is computed as $1 - \sum_{i,t} (n_{it} - p_{it})^2 / \sum_{i,t} (n_{it} - \overline{n})^2$, where p_{it} is our estimate for $\Pr[n_{it} = 1]$.

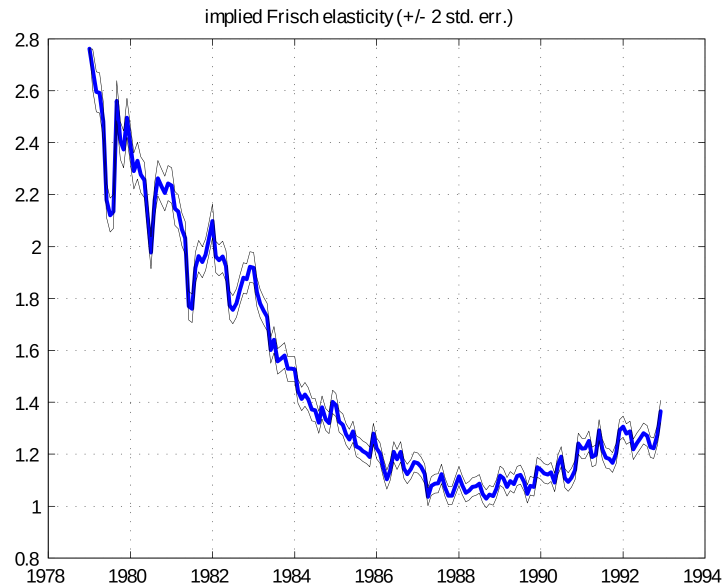

Figure 3: **The Aggregate Elasticity of Labor Supply, measured at each date.** The figure represents our point estimate, and a 95% confidence interval around it.

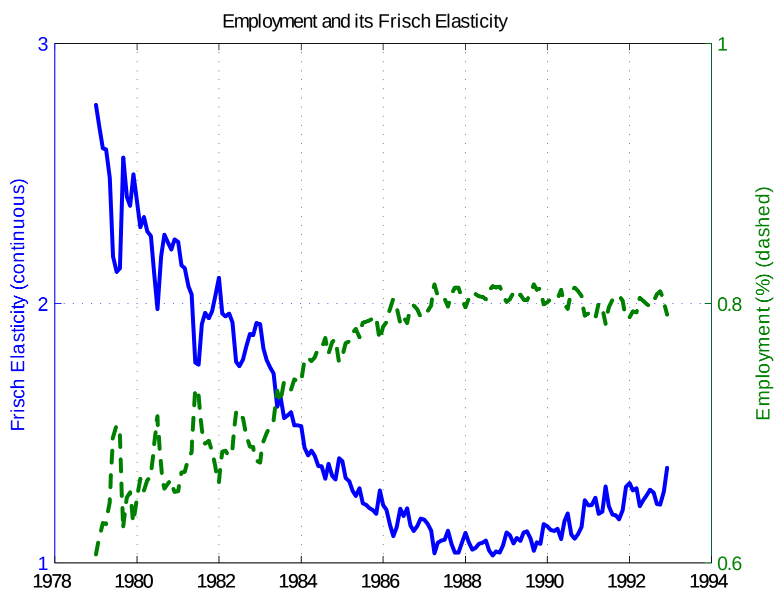

Figure 4: **Left scale**: the Frisch elasticity over time (thick plain line). **Right scale**: employment over time for our NLSY sample (dashed).

sample, evaluated at the mean of c_{it} at that date, is 1.83, while the true Frisch elasticity is 1.27 at this date.

Our estimate for the elasticity exhibits large fluctuations over time. These fluctuations are highly negatively correlated with aggregate employment. This is not surprising: if the density g does not decrease too fast, we would expect fluctuations in employment (the denominator in our measure of the elasticity) to make the elasticity countercyclical. Indeed, most of the hazard rates of the usual distributions are downward-sloping,[20] as is the one that we estimate below (Figure 5).

This negative correlation between the Frisch elasticity and employment has two facets. The first one is the life-cycle component, which dominates Figure 4. It is well known that young workers are more cyclical. For instance, Gomme, Rogerson, Rupert and Wright (2004) note that teenagers and young adults have more volatile employment, and are more elastic, than prime age workers. The second facet of this negative correlation is the business cycle, which we discuss in the next section.

Constructing labor supply schedules and hazard rates

Our estimation method allows us to construct the entire labor supply schedule, which is useful to understand the model. Indeed, we can compute a counterfactual latent index c_{it} for different hypothetical realizations of the aggregate wage rate w_t (the marginal product of an efficiency unit of labor). This amounts to varying b_t, the time effect that captures these fluctuations in w_t. Remember that

$$c_{it} \equiv \frac{b_t + a_i + \mathbf{x}_{it}\gamma - \left(b_t^R + a_i^R + \mathbf{y}_{it}\eta\right)}{\sigma},$$

and predicted employment for a given wage is just the sum of employment probabilities $\Phi\left(c_{it}\right)$ evaluated at a given b_t.

Figure 5 depicts the aggregate labor supply schedule, i.e. aggregate employment at different wage rates, at a given date t (we choose the median date of our sample, i.e. July 1985). The right panel represents the hazard rate of the distribution, evaluated at each wage rate. The vertical line represents the realized wage rate w_t at this date. The predicted employment at this date

[20] All distributions which are log-concave have downward-sloping hazard. Formally, if F is log-concave, then the hazard rate $f/F = d(\log F)$ is decreasing since $d^2 \log F < 0$.

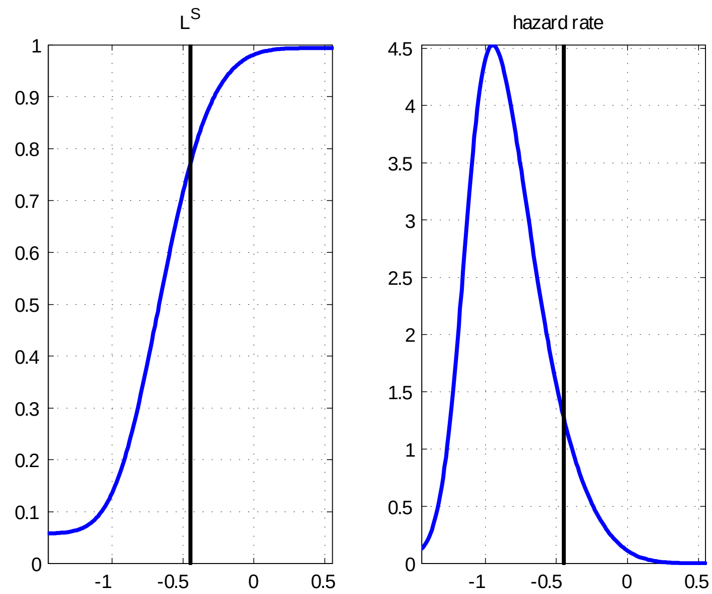

Figure 5: **Aggregate Labor Supply Curve and Hazard Rate.** *Left panel*: aggregate employment as a function of the aggregate wage rate w_t. *Right panel*: the corresponding hazard rate, i.e. the Frisch elasticity. The vertical line represents the realized wage rate in July 1985.

can be read off the graph, on the left panel, at the intersection of the cumulative distribution and the vertical line. Similarly, the Frisch elasticity can be read off the right panel, at the intersection of the estimated hazard rate and the vertical line. The negative correlation between employment and the Frisch elasticity can be seen in this figure: as employment decreases, the cutoff shifts to the left, and the hazard rate (i.e. the elasticity) consequently increases. Our formula for the Frisch elasticity is naturally decomposed across subpopulations. That is, we can compute an elasticity for each subsample of our population, and – weighting these elasticities by average employment probabilities – recover the aggregate elasticity. Figures 6 and 7 perform this exercise by separating our sample by gender or by education. These graphs show that women and the less-educated work less and are more elastic respectively than men and the more educated. This is a standard result in the labor literature (e.g. Heckman 1993, and Gomme, Rogerson, Rupert and Wright 2004). The Frisch elasticity can be read off on the hazard rate at the threshold value. For the median date (July 1985), the Frisch elasticity for

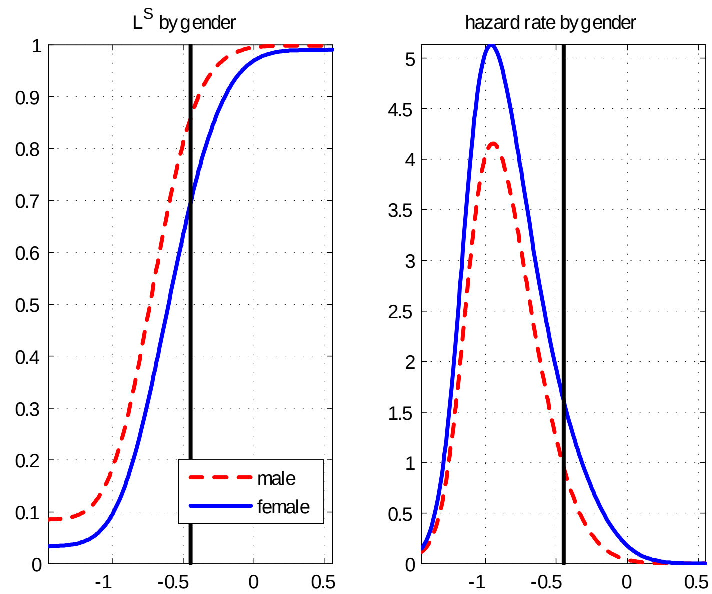

Figure 6: **Comparison of employment and elasticities by gender**: for both men and women, employment (resp. the Frisch elasticity) can be read on the y-axis, at the intersection of the CDF (resp. the hazard rate curve) and the vertical line which represents the realized wage rate in July 1985.

female is 1.62 and the Frisch elasticity for male is 0.96. The averages over the whole sample are respectively 1.82 and 1.19. The Frisch elasticities for college graduates, high-school graduates, and high-school dropouts at the median date are 1.13, 1.22 and 2.05. Similarly, the elasticities that we estimate for women with young children or people with health problems are above 2. In each of these decompositions of population across observable characteristics, we can check that the aggregate elasticity at this date is a weighted average of the different groups' elasticities, where the weights are the relative employment shares.

Our results are consistent with an estimate of the aggregate Frisch elasticity around 1 for prime-age workers (i.e. at the end of our sample). This estimate takes only into account the extensive margin. To give a definite answer regarding the Frisch elasticity of the whole population, we would need a sample representative of the US population. We have no workers older than 35 years old. But our results support the view that estimates based on samples of prime age males are highly misleading because they do not incorporate the workers which are

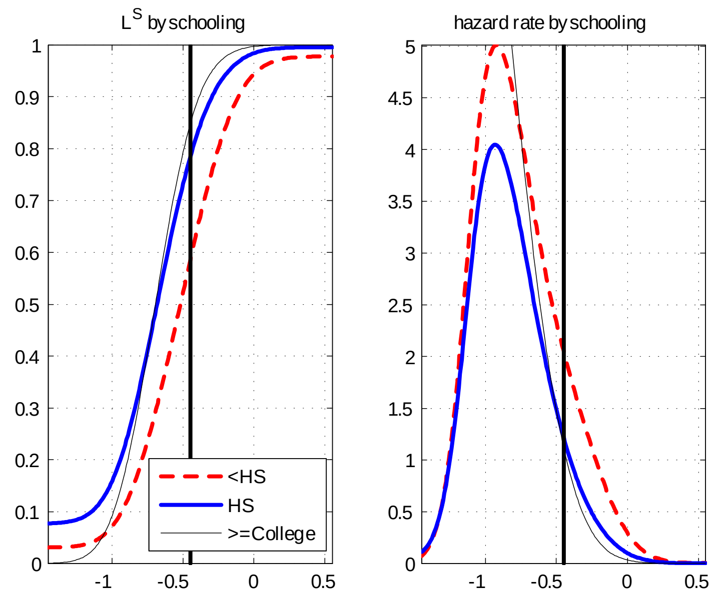

Figure 7: **Comparison of employment and elasticities by schooling level**. <HS = dropout, HS = high school degree and no BA, College= BA or higher. See notes to Figure 6.

elastic to aggregate fluctuations (women and young people). Moreover, Figure 3 shows that the elasticity appears to stabilize as workers enter their 30s, which suggests that one is a reasonable lower bound for the aggregate elasticity.

C. **Time-varying aggregate elasticity**

Our estimates also show that the labor supply elasticity is significantly countercyclical. To show this, we need to separate the effect of age and macroeconomic conditions. We use the differences in age between the respondents: more precisely, we first compute aggregate employment and the Frisch elasticity for the population of workers aged between 28 and 30, over the period 1987-1992: for each date t between 1987 and 1992, we include an individual i in our computation of the Frisch elasticity and employment only if his or her current age is 28, 29 or 30. Figure 8 presents the results. This figure demonstrates that the Frisch elasticity displays fluctuations even when age is held constant. The Frisch elasticity is higher during the NBER recession of 1990-1991 and the ensuing 'jobless recovery' than in earlier years.

Of course, the changing composition of the NLSY is a problem for us, since we would like

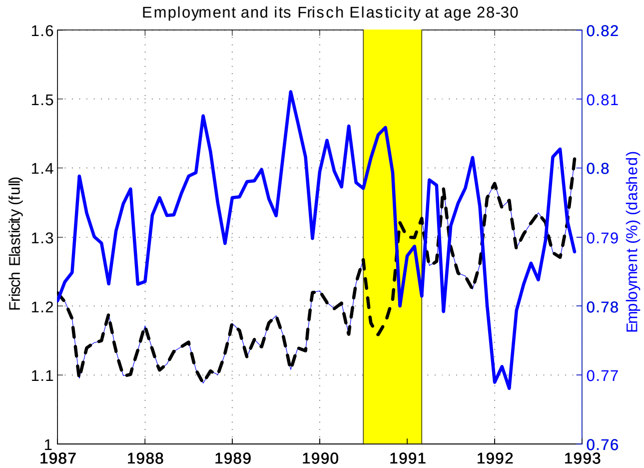

Figure 8: **Employment and the estimated Frisch elasticity for respondents aged 28 to 30.** This figure displays the Frisch elasticity and employment, holding age constant (see text). The shaded area represents the 1991 recession, according the NBER. As employment (full blue line) is low during the recession and jobless recovery, the elasticity (dashed and black) is higher.

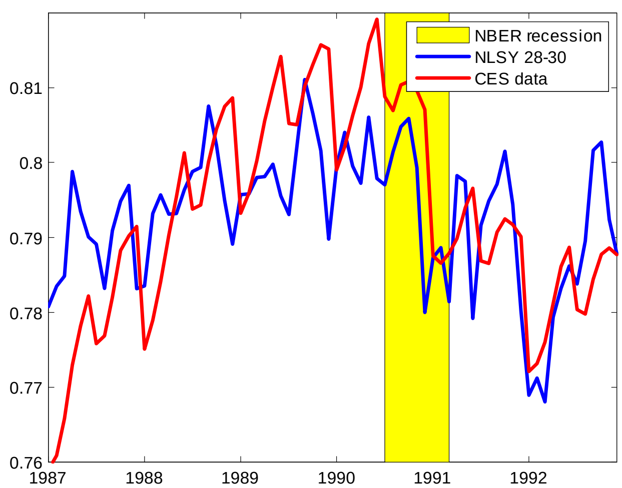

Figure 9: **Employment for respondents aged 28 to 30 and the CES number of employees.** The CES series is normalized to have the same mean as the NLSY series.

to have a representative sample of the entire US population. In earlier work, we used the PSID which is better in this respect. However, the NLSY is unique in providing monthly information on employment. In this respect, it is interesting to see that the NLSY results match standard national statistics. Figure 9 compares the ratio of employment to population in our sample for respondents aged 28 to 30 and the total number of employees according to the CES survey of the BLS. (The two series are normalized to have the same average.) The high correlation, in particular during the recession of 1990-1991 suggests that our procedure to keep age constant works well.

The next two pictures study the 1982 recession, during which the sample respondents were younger. We find a similar pattern using people aged 22 to 24. In that case their employment rate was significantly more cyclical than the CES (consistent with the fact that young people are more cyclical), but it remains highly correlated with it. Moreover, the Frisch elasticity appears to be countercyclical. Table 5 below reports the correlations of employment and the Frisch elasticity, holding age constant, for each age group. These correlations are all significantly negative.

We find these results interesting. At the very least, they suggest that the response of the economy to fundamental shocks may be state dependent. A natural conjecture is that this time-varying elasticity can create the appearance of a 'wedge' in the first order condition for leisure.[21] This wedge has been documented by many authors (starting perhaps with Hall 1997). In appendix B, we show that the model-implied wedge is $-\frac{1}{2}\left(\frac{1}{\phi_t}-\frac{1}{\phi^*}\right)\widehat{N}_t$. We derive the conditions for our model to generate the appearance of a wedge that is procyclical and as volatile as it is in the data. The condition has to do with the curvature (second derivative) of the relation between elasticity and hours. Our tentative conclusion based on numerical examples and measurement in the data of the expression $-\frac{1}{2}\left(\frac{1}{\phi_t}-\frac{1}{\phi^*}\right)\widehat{N}_t$ is that the magnitude of the wedge implied by the model is too small to explain the data wedge.

D. Heterogeneity and the aggregate elasticity

As argued in section 2, measuring realistically the heterogeneity in the data is crucial to obtain the aggregate elasticity of labor supply. This section shows how different specifications

[21] We thank Yongsung Chang for this suggestion.

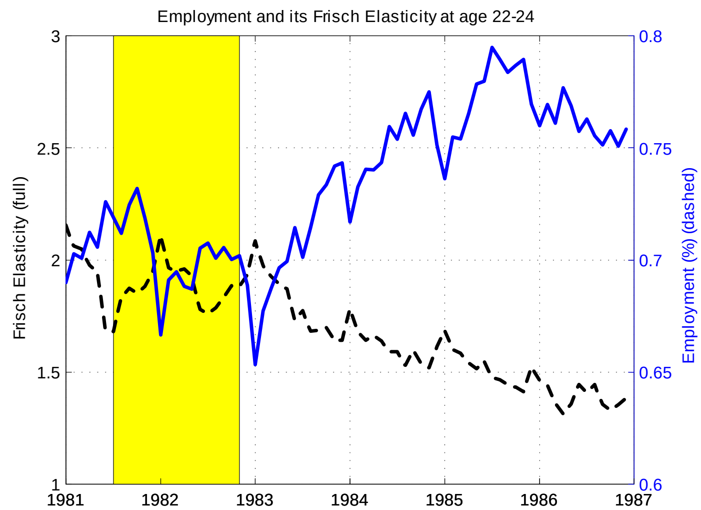

Figure 10: **Employment and the estimated Frisch elasticity for respondents aged 22 to 24.** This figure displays a measure of the elasticity at a constant age (Thus the respondents included in these aggregate measures changes accordingly.) The shaded area represents the 1982 recession, according the NBER. As employment (full blue line) is low during the recession, the elasticity (dashed black line) is higher.

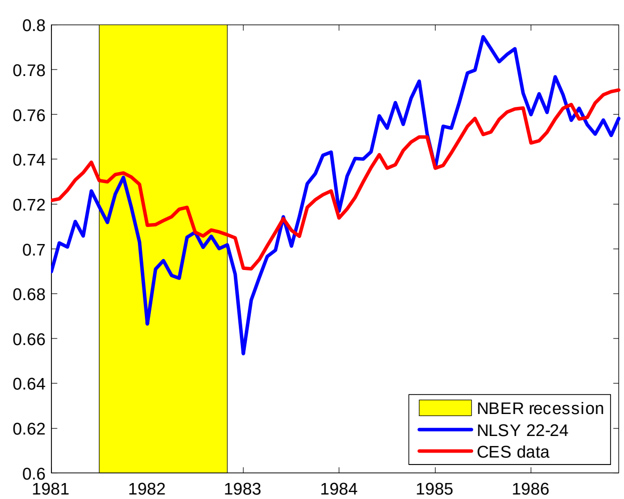

Figure 11: **Employment for respondents aged 22 to 24 and the CES number of employees.** The CES series is normalized to have the same mean as the NLSY series.

for the heterogeneity lead to different estimates for the elasticity.

For instance, consider the important work by Chang and Kim (2008). These authors study a model similar to ours, but with incomplete markets. Their paper, however, assumes that heterogeneity is only due to idiosyncratic wage shocks. But to compute the aggregate elasticity, one should ideally incorporate all the heterogeneity, i.e. not only the wage, but also other observables (education, experience, gender, ...) as well as unobserved heterogeneity, and preference shocks.Our specification has all three: key observables are captured in the vectors $\mathbf{x}$ and $\mathbf{y}$, unobservable heterogeneity is captured through the fixed effects, and the preference shock e_{it} has a significant variance.

Table 7 gives the results obtained by estimating some restricted specifications of our model. When preference shocks are assumed away, the elasticity is much higher. Intuitively, there is much less heterogeneity. When fixed effects are assumed away, the elasticity becomes enormous, since workers are nearly homogeneous. Removing the observables from either the participation decision, or the wage equation, also changes substantially the results. Hence, these assumptions appear to be very important quantitatively, which justifies our effort to develop a micro model which accounts for the large heterogeneity in the data.

	Aggregate Elasticity
Benchmark	1.27
No i.i.d. preference shock ($\sigma_e = 0$)	2.20
No fixed effects ($a_i = a_i^R = 0$)	8.73
No variables $\mathbf{x}$	1.56
No variables $\mathbf{y}$	1.05
No i.i.d. preference shocks and no $\mathbf{y}$ ($\sigma_e = 0$)	2.26

Table 7: Elasticity implied by various assumptions on the heterogeneity

E. **The Marginal Workers**

A large part of heterogeneity is not attributable to observable characteristics such as age, gender or schooling. This is what motivated us to use fixed effects and to go beyond simple decompositions by demographic group (See appendix C for a decomposition by groups using

CPS data.) In this subsection, we use the latent index c_{it}, which determines participation, to identify "marginal workers". More precisely the combination of fixed effects and observable covariates $\mathbf{x}_{it}, \mathbf{y}_{it}$ allows to identify the agent who is closest to being indifferent between working or not: the latent index c_{it} is nearly equal to zero for this agent. We define marginal workers as workers whose c_{it} is small enough, and we choose a threshold $|c_{it}| \leq 0.5$. This is equivalent to defining marginal workers at time t as the individuals for which the model predicts a probability of working higher than $0.305\% = \Phi(-0.5)$ and less than $0.695\% = \Phi(0.5)$.

Figure 12 plots the share of the population which is 'marginal' by this definition. This share falls steeply at the beginning of our sample before stabilizing around 12%. This is clearly in large part driven by age and experience effects. Figure 13 shows the evolving distribution of the index c_{it} over time. The vertical lines at 0 represent the marginal worker. In 1979, when agents are all young, they are nearly homogeneous, and most agents have c_{it} close to zero. As time goes by, c_{it} drifts the right, most agents are permanently employed, and the marginal workers are not any more the mode of the distribution. The figure shows that the distribution gets skewed, and that the variance first increases, probably as people enter the workforce at different ages. The model predicts that the elasticity of labor supply is high when the share of marginal workers is high. This is clearly true along the life cycle (compare figure 12 and figure 3), and table 5 shows this is also true along the business cycle: that table presents the correlations between employment rates, frisch elasticity, and the number of marginal workers holding age constant. Figure 14 illustrates this correlation for the group of 25-year old.

Age group	22	23	24	25	26	27	28	29	30
Corr(N,Frisch)	-0.45	-0.70	-0.79	-0.88	-0.72	-0.63	-0.54	-0.37	-0.62
Corr(N,Marg)	-0.23	-0.48	-0.72	-0.88	-0.73	-0.48	-0.73	-0.53	-0.24
Corr(Marg,Frisch)	0.90	0.90	0.95	0.98	0.97	0.92	0.74	0.37	-0.11

Table 5: For each group of constant age, this table reports in each column the correlation over time between the employment rate, the Frisch elasticity, and the share of marginal workers (Marg), over the business cycle. For instance, the sample of age group 22 runs from 1979 to 1985.

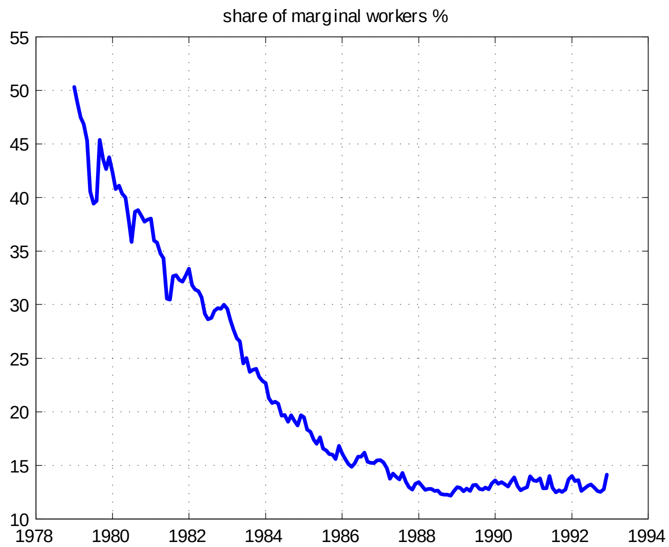

Figure 12: **Share of population who are "marginal workers"**. Marginal workers are defined by $|c_{it}| \leq 0.5$.

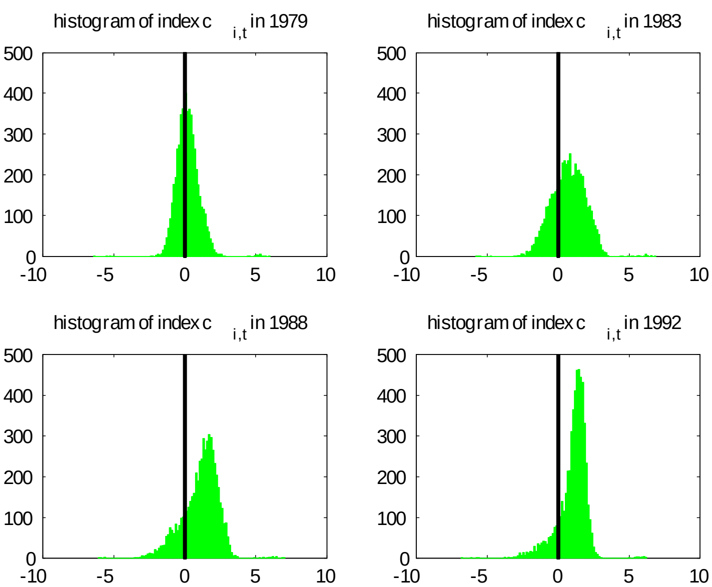

Figure 13: **The distribution of the index c_{it} at various dates**: 1979, 1983, 1988, 1992. The figure illustrates how the distribution fans out and becomes skewed over time.

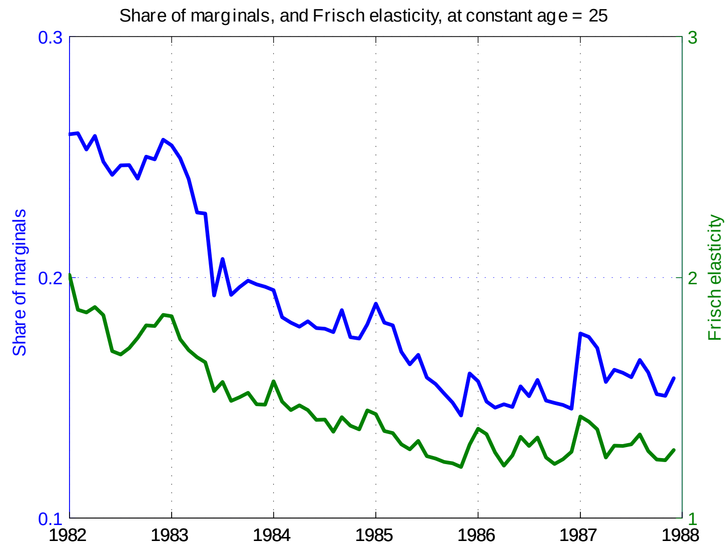

Figure 14: **Share of marginals and Frisch elasticity**, for the group of 25-year old people. We substitute cohorts to keep the age constant. The Frisch elasticity of this group is highly correlated with the share of marginals in this group.

Table 6 presents some summary statistics on the population of marginal workers at the end of our sample (i.e. for people about 30 years old). Marginal workers are less experienced, slightly less schooled, and are often women with young children. Marginal workers are more likely to feel constrained by health in their choice of work. Being a marginal worker is a persistent status, but not highly so: the probability of being marginal the next month if you are marginal this month is 0.945, while it is 0.017 if you were not marginal this month.

Means	whole population	marginal workers
experience (yrs)	10.15	7.58
schooling (yrs)	13.29	12.74
gender (male=1)	0.46	0.20
marital Male dummy	0.26	0.05
marital Female dummy	0.34	0.61
health limit dummy	0.07	0.15
woman with kid 0-2 dummy	0.10	0.32
woman with kid 3-6 dummy	0.14	0.24
woman with kid 7-14 dummy	0.13	0.17

Table 6: Summary statistics for the population of marginal workers and for the whole population in December 1991

We can then test the hypothesis that these marginal workers account for a large share of aggregate fluctuations. We decompose changes in aggregate employment into the sum of the changes of the marginal worker population and the non-marginal worker population:

$$\Delta\widetilde{N}_t = \Delta\widetilde{N}_t^{\text{marginal workers}} + \Delta\widetilde{N}_t^{\text{non marginal workers}},$$

and we measure the contribution of marginal workers to aggregate fluctuations as their share in the variance, i.e.:

$$\frac{Cov\left(\Delta\widetilde{N}_t, \Delta\widetilde{N}_t^{\text{marginal workers}}\right)}{Var\left(\Delta\widetilde{N}_t\right)}.$$

Over 1979-92, the group of marginal workers (who represents 21.8% of the population on average) accounts for 48.6% of aggregate fluctuations in employment. Marginal workers also account for 32.6% of transitions between employment and non-employment. The c_{it} we measure is only an estimate of the latent index, before current shocks are realized. Therefore the variance of transitory shocks – which is approximately as big as the variance of c_{it} – introduces some 'noise' and implies that our marginal workers do not account for 100% of the variance of aggregate employment.

As explained above, the degree of persistence of being marginal has no direct effect on the aggregate elasticity. Hence, these calculations about marginal workers are really a joint test of the indivisible labor model (with complete markets) and a particular process for tastes and abilities: fixed effect plus i.i.d. To test our joint hypothesis more precisely, we simulate data from the estimated model and apply the same procedure to define marginal workers. The model predicts a share of variance of aggregate employment accounted for by marginal workers of 45.8%. This is close to that obtained from the data (48.6%). The model predicts however that marginals do more transitions (41.5%) than they do in the data (32.6%).

5 Robustness

In this section, we investigate the robustness of our findings to several extensions. Specifically, we study how the results are affected when the idiosyncratic shocks are persistent rather than *i.i.d.*, we study the effect of measurement error, and we provide monte-carlo evidence of the performance of our estimator.

A. Persistence of idiosyncratic shocks

Our assumption that the shocks to wages and to labor disutility (u_{it} and e_{it}) are independent and identically distributed over time is at odds with the empirical literature on wage processes, which shows that shocks are highly persistent. One may wonder how our results are affected by this assumption.[22]

It is important to note that in our model, the persistence of shocks is completely *irrelevant*. This is because the labor supply is given by $\widetilde{N}_t = G(\log x_t^*)$, where G is the *unconditional* distribution of productivity, tastes and wealth (see equation 2.3). With complete markets, only ex-post heterogeneity matters, and risk is irrelevant.

An altogether different issue is that assuming i.i.d. shocks may bias our estimates. It is difficult to answer this question analytically, so we resort to a numerical experiment. Using our parameter estimates, we simulate artificial data from our model, except that we assume now

[22] As explained in Section 3, this assumption is driven by the impossibility of estimating a nonlinear panel data model with persistent unobserved shocks and fixed effects, when T is large.

that u_{it} and e_{it} follow AR(1) processes:

$$\begin{aligned} u_{it+1} &= \rho_u u_{it} + \varepsilon_{u,i,t+1}, \\ e_{it+1} &= \rho_e e_{it} + \varepsilon_{e,i,t+1}, \end{aligned}$$

where $(\varepsilon_{u,i,t+1}, \varepsilon_{e,i,t+1})$ is an i.i.d. and normally distributed vector with mean $(0,0)$ and covariance matrix Σ. We run our estimator - which presumes that shocks are i.i.d. - on these artificial data and compute the implied aggregate elasticity of labor supply.

More precisely, we pick a value for ρ_u and ρ_e, and then pick the covariance matrix Σ to obtain the same unconditional covariance matrix of (u_{it}, e_{it}) as in our i.i.d. case. This ensures that we have the same unconditional distribution, but now the conditional distribution is very different.[23]

Table 8 reports the effect of varying the persistence on our estimated values for the elasticity of labor supply.[24] The central result is that the estimated elasticity of labor supply is hardly affected by the persistence parameter. As ρ goes towards 1, there is a slight negative bias, which suggests that our benchmark results may slightly underestimate the aggregate elasticity.

	Truth	Estimated			
		$\rho = 0$	$\rho = .9$	$\rho = .98$	$\rho = .995$
Elasticity	1.27	1.253	1.223	1.185	1.174

Table 8: Effect of persistence on the estimated elasticity of labor supply.

Intuitively, two opposite effects arise. First, for people who are nearly indifferent between working and not, the variance of the shock is now smaller, which tends to make the elasticity larger. Second, because of the persistence, people typically "drift away" from the point where they are indifferent between working and not working. This means there are fewer marginal agents and a lower elasticity. Overall, the aggregate elasticity is not substantially affected, since the bias is less than 10%.

[23] Formally, we pick $\Sigma_{u,u} = \sigma_u^2(1-\rho_u^2)$, $\Sigma_{v,v} = \sigma_v^2(1-\rho_v^2)$, and $\Sigma_{u,v} = \sigma_{u,v}(1-\rho_u\rho_v)$, where the σ refer to our benchmark estimates.

[24] Here, we assume that $\rho_u = \rho_e$ for clarity. Having different persistence parameters does not affect these results.

B. Role of measurement error

An important question is how measurement error impacts our results. Intuitively, measurement error implies that we overestimate the amount of heterogeneity in the data, and hence typically we underestimate the aggregate elasticity of labor supply.

To assess the importance of this effect, we simulate artificial data from our model (using the estimated parameter values). We next add some measurement error to this artificial data. Finally, we use our estimation procedure and report in Table 9 the aggregate elasticity that we would infer from these data.

There are different possible types of measurement error. Row 1 shows our benchmark results, which assume no measurement error. Rows 2 and 3 show the effect of measurement error in the wage only: an i.i.d. normal, classical measurement error is added to the log of the wage, with standard deviation 5% in row 2 and 10% in row 3.. Rows 4 (resp. row 5) show the effect of measurement error in the employment status. Here it is assumed that a 5% (resp 10%) share of the employed is misreported as non-employed, and similarly 5% (resp. 10%) of non-employed are reported as employed. Note that, as a result, total employment is reduced, since there are more employed than non-employed. As an alternative, rows 6 and 7 consider the case where the number of non-employed reported as employed is such that total employment is not affected.

	Measurement error (ME)	Aggregate Elasticity
1	None	1.27
2	ME in wages: 10%	1.27
3	ME in wages: 20%	1.27
4	ME in participation: 5%	1.21
5	ME in participation: 10%	1.18
6	ME: E→U 5%, and U→E to keep same total employment	1.03
7	ME: E→U 10%, and U→E to keep same total employment	0.85

Table 9: Effect of measurement error on the estimated aggregate elasticity of labor supply.

The table reveals first that measurement error in the wage does not affect our procedure.

Intuitively, this is because the key parameter σ is estimated from the ratio of the coefficients in the probit and the wage equation, and adding independent noise to the wage equation does not bias our estimates for the coefficients. Second, measurement error in the employment status does affect our results. Measurement error increases the size of the shocks that we estimate, suggesting that there is more heterogeneity than there is actually, and hennce reducing the measured elasticity. This is confirmed in rows 4 through 7. Conversely, this table suggests that our estimate for the elasticity is a lower bound, since we assumed that there was no measurement error.

C. Monte-carlo experiments

An econometric concern with our estimation procedure is that a nonlinear panel data estimator with fixed effects is in general biased. This is known as the incidental parameter problem. This is especially a problem in panel data model with small T, since the number of observations per individual is small, leading to a bias in the fixed effect which may contaminate the other parameter estimates. In our model, T is relatively large, so this problem would seem less acute. To check this, we run Monte-carlo simulations by simulating 100 artificial panel data sets from our model. We then estimate each panel data set and recover the parameters. Table 10 presents the average and standard deviation of parameter estimates, across the 100 data sets, together with the "true" parameters used to simulate artificial data. (We use our estimated parameter estimates as the "true" parameters.) The table reveals that the bias is indeed small, and that the variation across data sets is also small. Hence in our case, the incidental parameter problem does not appear to be very important.[25]

[25] The fixed effects are estimated with a small bias, but substantial imprecision, however this does not affect our conclusions, e.g. the aggregate elasticity.

	Name	Truth	Mean estimate	Std Dev
Variable **x**	experience	-0.0564	-0.0563	0.0024
	experience squared	0.0403	0.0401	0.0018
	schooling	0.0771	0.0771	0.0026
Variable **y**	married (men)	0.2145	0.2142	0.0043
	married (women)	0.1022	0.1020	0.0031
	health limit	0.0470	0.0469	0.0036
	youngest kid 0-2 years	0.1216	0.1213	0.0014
	youngest kid 3-6 years	-0.0026	-0.0026	0.0001
	youngest kid 7-14 years	0.0988	0.0988	0.0007
Shocks	σ_u	0.1240	0.1251	0.0017
	σ_{ue}	0.1230	0.1239	0.0009
	σ_e	0.1531	0.1522	0.0004
	σ	0.1763	0.1753	0.0030
Aggregate elasticity		1.27	1.29	0.0183

Table 10: Monte carlo results. We simulate 100 artificial data sets, run our estimator on each data set, and report the mean and std. dev. of estimates.

6 Conclusion

This paper makes two main points. First, some agents have a more cyclical labor supply. Second, the aggregate elasticity of labor supply is related to the homogeneity of the workforce at the margin, i.e. to the number of "marginal workers": these marginal workers are nearly indifferent between work and leisure at a given point in time, hence fluctuations in the aggregate wage drive them in and out of the workforce.

We develop an empirical framework to measure the elasticity implied by this heterogeneity. Our estimate for the Frisch elasticity of aggregate labor is 1.5 over our whole sample. This elasticity varies over the life cycle and over the business cycle: it is countercyclical. We find that the elasticity is driven by the number of 'marginal workers', which are disproportionately female, have children, have health problems, somewhat less schooling, and are less experienced.

These marginal workers change over time. They account for a significant share of aggregate fluctuations: over 1979-1992, 22% of agents account for about 49% of aggregate fluctuations in employment.

There are several directions in which to extend this work. First, it seems interesting to incorporate an intensive margin of labor supply for people who are working. In ongoing work, we show that our approach can be generalized to this setup; this adds an additional term to our elasticity. Second, it may seem important to improve the fit of the econometric model by allowing for persistent shocks or costs of job search (Altug and Miller 1998, Hyslop 1999). This would require us to drop the unrestricted shape of heterogeneity due to fixed effects, so we leave this for future research. Third, a limitation of our approach is that maximum likelihood estimation requires to make distributional assumptions on the unobserved stochastic heterogeneity (the iid shock), even if we identify permanent heterogeneity nonparametrically with fixed effects. Because we estimate idiosyncratic shocks to be large, the hazard rates that give the Frisch elasticity inherit the gaussian shape of the iid shock. It would be interesting to measure the distribution of agents' wages and reservation wages around the marginal worker with less stringent distributional assumptions on this shock.

References

1. Alesina, Alberto, Edward Glaeser, and Bruce Sacerdote, 2005. "Work and Leisure in the US and Europe: Why So Different?". NBER Macroeconomics Annual.

2. Altonji, Joseph G. 1986. "Intertemporal Substitution in Labor Supply: Evidence from Micro Data," December 1984, *Journal of Political Economy*, Vol. 94, No. 3, Part 2, June: S176-S215.

3. Altug, Sumru and Robert A. Miller. 1998. "The Effect of Work Experience on Female Wages and Labour Supply," *Review of Economic Studies*, vol. 65(1), January: 45-85.

4. Amemiya, Takeshi. 1985. *Advanced Econometrics*. Blackwell eds.

5. Benhabib, Jess and Rogerson, Richard and Wright, Randall, 1991. "Homework in Macroeconomics: Household Production and Aggregate Fluctuations," *Journal of Political Economy*, vol. 99(6), December: 1166-87.

6. Bils, Mark, Yongsung Chang and Sun-Bin Kim, 2007. "Comparative Advantage in Cyclical Unemployment". Mimeo, University of Rochester.

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

8. Chang, Yongsung and Sun-Bin Kim. 2006b. "Heterogeneity and Aggregation: Implications for Labor-Market Fluctuations". Forthcoming, American Economic Review.

9. Chari, V.V., Patrick Kehoe and Ellen McGrattan. 2006. "Business Cycle Accounting". Econometrica, forthcoming.

10. Cho, Jank-Ok and Richard Rogerson. 1988. "Family Labor Supply and Aggregate Fluctuations". *Journal of Monetary Economics*, 21:233-245.

11. Cho, Jang-Ok. 1995. "Ex post heterogeneity and the business cycle," *Journal of Economic Dynamics and Control*, Elsevier, vol. 19(3), pages 533-551, April.

12. Clark, K.B. and Lawrence Summers. 1981. "Demographic Differences in Cyclical Employment Variation". Journal of Human Resources, 16:61-79.

13. Coleman, Thomas Sedgwick. 1984. "Essays on Aggregate Labor Markets Business Cycle Fluctuations", unpublished dissertation, University of Chicago.

14. Gomme, Paul, Richard Rogerson, Peter Rupert and Randall Wright. 2004. "The Business Cycle and the Life Cycle." *NBER Macroeconomics Annual*, 415-461.

15. Greene, William. 2001 "Estimating Econometric Models with Fixed Effects." working paper, New York University.

16. Hadegorn, Marcus and Iourii Manovskii. 2005. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies Revisited". working paper, University of Pennsylvania.

17. Haefke, Christian and Michael Reiter, 2006. "Endogenous Labor Market Participation and the Business Cycle", Mimeo, Pompeu Fabra.

18. Hall, Robert E, 1997. "Macroeconomic Fluctuations and the Allocation of Time," *Journal of Labor Economics*, vol. 15(1), January: S223-S250.

19. Hall, Robert E., 2006. "Sources and Mechanisms of Cyclical Fluctuations in the Labor Market", working paper, Stanford University.

20. Hansen, Gary. 1985. "Indivisible Labor and the Business Cycle." *Journal of Monetary Economics*, 16 (3) November: 309-327.

21. Heckman, James J. 1993. "What has been Learned about Labor Supply in the Past Twenty Years?" *American Economic Review*, 83 (2), May: 116-121.

22. Heckman, James J., 1979. "Sample Selection Bias as a Specification Error," *Econometrica*, vol. 47(1), January: 153-61.

23. Heckman, James J. and Thomas MaCurdy. 1980. "A Life Cycle Model of Female Labour Supply." *Review of Economic Studies*, vol. 47(1), January: 47-74.

24. Heckman, James J. and Thomas MaCurdy, 1982. "Corrigendum on a Life Cycle Model of Female Labour Supply," *Review of Economic Studies*, vol. 49(4), October: 659-60.

25. Heckman, James J., Lance Lochner and Petra Todd. 2003. "Fifty Years of Mincer Earnings Regressions." Forthcoming, *Handbook of Education Economics*.

26. Hyslop, Dean. 1999. "State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women." *Econometrica*, Vol. 67 No.6, November: 1255-1294.

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

28. Kimmel, Jean and Thomas J. Kniesner. 1998. "New evidence on labor supply: Employment versus hours elasticities by sex and marital status." *Journal of Monetary Economics*, vol. 42(2), pages 289-301, July.

29. King, Robert G. and Sergio T. Rebelo. 1999. "Resuscitating Real Business Cycles." in: J. B. Taylor and M. Woodford (ed.), *Handbook of Macroeconomics*, Elsevier, edition 1, volume 1, chapter 14, 927-1007.

30. Kydland, Finn. 1984. "Labor Force Heterogeneity and the Business Cycle". Carnegie-Rochester Conference Series on Public Policy. 21:173-208.

31. Ljungqvist, Lars and Thomas Sargent. 2006. "Do Taxes Explain European Employment? Indivisible Labor, Human Capital, Lotteries, and Savings", *NBER macroeconomics annual*.

32. Lucas, Robert E. and Leonard Rapping. 1969. "Real Wages, Employment and Inflation." *Journal of Political Economy* 77, 721-754.

33. MaCurdy, Thomas E. 1981. " An Empirical Model of Labor Supply in a Life-Cycle Setting." *Journal of Political Economy.* Vol. 89, No. 6, December: 1059-1085

34. Manovskii, Iourii and Marcus Hagedorn. 2006. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies Revisited." Working paper, University of Pennsylvania.

35. Mulligan, Casey B. 2001. "Aggregate Implications of Indivisible Labor." *Advances in Macroeconomics* 1 (1) www.bepress.com/bejm/advances/vol1/iss1/art4 (also NBER #8159), March.

36. Mulligan, Casey B. 2002. "A Century of Labor-Leisure Distortions" NBER working paper #8774.

37. Pencavel, John. 1986. "Labor Supply of Men." in: *Handbook of Labor Economics.* Ashenfelter and Layard (eds.) North Holland.

38. Prescott, Edward C. 2002. "Prosperity and Depression." *American Economic Review* 92. May: 1-15.

39. Prescott, Edward C. 2006. "Nobel Lecture: The Transformation of Macroeconomic Policy and Research." *Journal of Political Economy.* Vol. 114, No. 2, 203-235.

40. Rogerson, Richard. 1988. "Indisible Labor, Lotteries and Equilibrium." *Journal of Monetary Economics.* 21(1) January: 1-16.

41. Rogerson, Richard and Peter Rupert. 1993. "On testing the intertemporal substitution theory of labor supply." *Journal of Economic Dynamics and Control* 17 (1993) 37-50.

42. Rogerson, Richard and Johanna Wallenius, 2007. "Micro and Macro elasticities in a Life Cycle Model with Taxes". Mimeo, ASU.

43. Rotemberg, Julio and Michael Woodford . 1997. "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy", *NBER Macroeconomics Annual*, 297-346.

44. Shimer, Robert. 2005. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies." *American Economic Review*, 95(1): 25-49.

45. Solon, Gary, Robert Barsky and Jonathan A Parker. 1994. "Measuring the Cyclicality of Real Wages: How Important Is Composition Bias?" *Quarterly Journal of Economics*, vol. 109(1), February: 1-25

46. Townsend, Robert M. , 1994. "Risk and Insurance in Village India", Econometrica 62(3):539-591

47. Woodford, Michael. 2003. *Interest and Prices.* Princeton University Press.

7 Appendix

A. Data Description

We use panel data from the National Longitudinal Survey of Youth 1979 (NLSY 79). The sample consists of a cohort of men and women, born between 1957 and 1964, surveyed annually from 1979 until 1994, and then every other year until now. These data provide a detailed account of each individual's work history, including precise dates for each employment spell, with the associated wage rate and hours worked. In addition, detailed demographic information is presented. This high frequency dataset is well suited to our purpose (as opposed to the PSID which is annual). We use monthly data drawn from the NLSY, for which the indivisibility seems to be a reasonable approximation.

Our sample only excludes the military supplements and the supplements for poor whites.[26] We analyze the period between January 1979 and December 1992: hence $T = 168$ months and $N = 5571$ agents. We use sampling weights as provided by the NLSY.

Our employment variable is constructed from weekly labor force status (employed, unemployed, or out of the labor force). Monthly labor force status is set as "employed" if the respondent was employed in any week of this month. Labor force status is interpreted as $n_{it} = 1$ if employed, and $n_{it} = 0$ if unemployed or out of the labor force. We balance our sample according to this employment measure.

Data on wages w_{it} come from information on hourly wage rates for up to five jobs for each survey year. Matching this information with labor force status is sometimes difficult and results in missing wage data (w_{it} is missing for approximately 15% of observations with $n_{it} = 1$). We had problems with the wage data after 1993 (when the survey methodology changed for these questions), which is why we restricted ourselves to 1979-1992. In the case of simultaneous jobs, we weight wage rates by average hours worked at each job. Information on hours worked is not used otherwise.

The construction of most observable covariates $\mathbf{x}_{it}, \mathbf{y}_{it}$ is straightforward from the data that

[26] That is, we include the following subsamples: cross male white, cross male white poor, cross male black, cross male hispanic, cross female white, cross female white poor, cross female black, cross female Hispanic, supplement male black, supplement male Hispanic, supplement female black, supplement female hispanic.

are available online. We constructed experience by summing cumulated months worked at each point in time. (Hence, it is a true experience variable, not age minus schooling minus 6.) The highest grade achieved at school is given in the dataset. The age of a respondent's youngest kid can be constructed accurately from the data.

Our empirical specification of wages and participation thus conforms to the general practice in labor economics. For instance, in their study on married women, Heckman and MaCurdy (1980, 1982) have a participation equation with the number of children and the number of children less than six years old, family income excluding the wife's earnings, the wife's age, her husband's hours unemployed, and whether he is retired or disabled. Hyslop (1999) includes race, age of youngest child, nonlabor income, marital status as well.

B. **Time-varying aggregate elasticity and the labor wedge**

Hall (1997) noted one failure of the labor supply model: the marginal rate of substitution between consumption and leisure and the marginal product of labor are only weakly correlated. There is a time-varying wedge in this first-order condition. To illustrate this, consider the case of a representative agent with utility function:

$$U = \log C_t - B\frac{N_t^{1+\frac{1}{\phi}}}{1+\frac{1}{\phi}},$$

where ϕ is the Frisch elasticity of labor supply and B is a constant. The first-order condition is $BC_tN_t^{\frac{1}{\phi}} = w_t = \alpha Y_t/N_t$, where the last equality assumes a Cobb-Douglas production function so that the marginal product of labor is proportional to the average product of labor. The wedge ξ_t is the error in this FOC: $\xi_t = BC_tN_t^{\frac{1}{\phi}}/(\alpha Y_t/N_t)$. For a given value of ϕ, we can measure this wedge (in log and HP-filtered): it is highly procyclical and roughly as volatile as hours. (I assume an elasticity equal to 1.5, consistent with the average estimated Frisch elasticity in our sample.) Figure 15 illustrates this puzzle by plotting the wedge and hours (both in log and HP filtered). This puzzle has attracted a lot of attention in the business cycle literature (e.g. Chari, Kehoe and McGrattan (2006), Mulligan (2002)).

Our model suggests a potential resolution of this puzzle: the elasticity of labor supply may be time-varying, which can create the appearance of a wedge if the researchers fits a model with a constant Frisch elasticity as above. We now derive an expression for the wedge implied

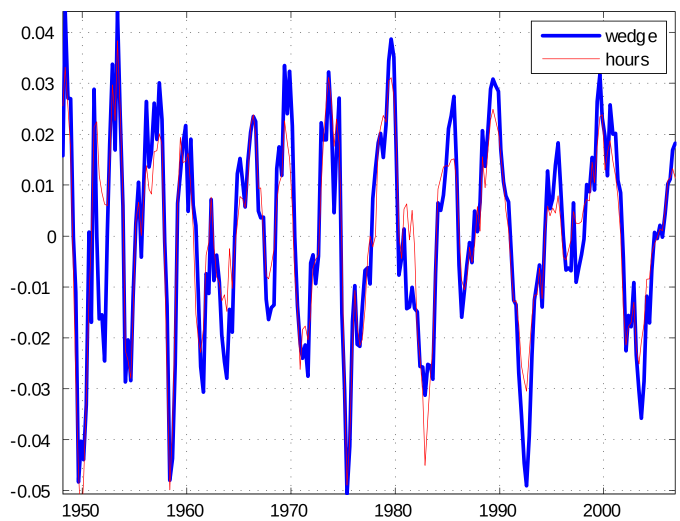

Figure 15: **HP filtered hours (thin, red) and HP filtered wedge ξ_t (thick, blue) in the first-order condition for leisure**. This computation assumes a utility function $\log C_t - BN_t^{1+1/\phi}/(1+1/\phi)$. The MPL is measured as average productivity, i.e. we assume a Cobb-Douglas production function.

by the model, and the conditions under which it will be procyclical. Our model yields a per period utility function for the representative agent, $U(C_t) - V(N_t)$ as explained in section 2. Assuming $u(C_t) = \log C_t$, so that the marginal utility of aggregate consumption is correctly specified, we have the following first order condition: $C_t V'(N_t) = w_t$. A researcher who measures the wedge as $\xi_t = C_t N_t^{\frac{1}{\phi}}/w_t$, would find in our economy a value $\xi_t = N_t^{1/\phi}/V'(N_t)$ given that $C_t V'(N_t) = w_t$. If V is not linear in log, ξ_t may be time-varying. It is interesting to approximate this expression. To take into account the possibility of a time-varying elasticity, we start with the second-order approximation to $\log V'(N_t)$ around the steady-state of hours:

$$\log V'(N_t) \simeq \log V'(N^*) + g(N^*)(N_t - N^*) + \frac{1}{2} g'(N^*)(N_t - N^*)^2, \tag{7.1}$$

where $g(N) = V''(N)/V'(N)$. To measure $g'(N^*)$, we now use the fact that we have an estimate of ϕ_t, which satisfies:

$$\frac{1}{\phi_t} = \frac{V''(N_t)N_t}{V'(N_t)} = N_t g(N_t).$$

Hence, writing a first-order approximation[27] for $g(N_t)$:

$$g(N_t) \simeq g(N^*) + g'(N^*)(N_t - N^*),$$

we obtain the following formula for $g'(N^*)$, given that $1/\phi^* = N^* g(N^*)$:

$$g'(N^*) \simeq \frac{g(N_t) - g(N^*)}{N_t - N^*} = \frac{\frac{1}{\phi_t N_t} - \frac{1}{\phi^* N^*}}{N_t - N^*},$$

and thus plugging this into the approximation (7.1), and denoting $\widehat{N}_t = (N_t - N^*)/N^*$:

$$\log \frac{V'(N_t)}{V'(N^*)} = \frac{1}{\phi^*}\widehat{N}_t + \frac{1}{2} \frac{\frac{1}{\phi_t N_t} - \frac{1}{\phi^* N^*}}{N_t - N^*} {N^*}^2 \widehat{N}_t^{\,2},$$

which after some simplifications yields:

$$\begin{aligned} \log \frac{V'(N_t)}{V'(N^*)} &\simeq \frac{1}{\phi^*}\widehat{N}_t + \frac{1}{2}\left(\frac{1}{\phi_t}\frac{1}{1+\widehat{N}_t} - \frac{1}{\phi^*}\right)\widehat{N}_t, \\ &\simeq \frac{1}{\phi^*}\widehat{N}_t + \frac{1}{2}\left(\frac{1}{\phi_t} - \frac{1}{\phi^*}\right)\widehat{N}_t. \end{aligned}$$

[27] A first-order approximation is sufficient here since we wish only to approximate $g'(N^*)$, the second order term in (6.1).

Hence, the wedge that a researcher would infer from our model if he used the correct average elasticity ϕ^*, but did not take into account the fact that the elasticity varies over time, is:

$$\begin{aligned}\log \xi_t &= \frac{1}{\phi^*}\log N_t - \log V'(N_t) \\ &\simeq \text{constant} + \frac{1}{\phi^*}\widehat{N}_t - \frac{1}{\phi^*}\widehat{N}_t - \frac{1}{2}\left(\frac{1}{\phi_t} - \frac{1}{\phi^*}\right)\widehat{N}_t \\ &\simeq \text{constant } - \frac{1}{2}\left(\frac{1}{\phi_t} - \frac{1}{\phi^*}\right)\widehat{N}_t.\end{aligned}$$

In the data, the wedge is procyclical and volatile. To replicate this feature of the data, we would need to have $Cov(\log \xi_t, \widehat{N}_t) > 0$ i.e. $Cov\left(\left(\frac{1}{\phi_t} - \frac{1}{\phi^*}\right)\widehat{N}_t, \widehat{N}_t\right) < 0$.

In this model, the elasticity depends on the hazard rate of G evaluated at the cutoff, and employment is the cdf G evaluated at the cutoff. If these functions are monotonic, we have a well-defined reduced form relation $1/\phi_t = h(\widehat{N}_t)$ along the equilibrium. (More precisely, $N_t = G(\log x_t^*)$ and $1/\phi_t = g(\log x_t^*)/G(\log x_t^*)$, hence $1/\phi_t = g(G^{-1}(N_t))/N_t \overset{def}{=} h_2(N_t) \overset{def}{=} h(\widehat{N}_t)$). Assuming that $\widehat{N}_t$ is normally distributed, we can use Stein's lemma:[28]

$$\begin{aligned}Cov\left(\left(\frac{1}{\phi_t} - \frac{1}{\phi^*}\right)\widehat{N}_t, \widehat{N}_t\right) &= Cov\left(\left(h(\widehat{N}_t) - h(0)\right)\widehat{N}_t, \widehat{N}_t\right) \\ &= E\left(h'(\widehat{N}_t)\widehat{N}_t + \left(h(\widehat{N}_t) - h(0)\right)\right) \times Var(\widehat{N}_t).\end{aligned}$$

Hence, the wedge is procyclical and large if and only if $E\left(h'(\widehat{N}_t)\widehat{N}_t + \left(h(\widehat{N}_t) - h(0)\right)\right)$ is negative and large in absolute value. Approximating the reduced form function h, we obtain, using $E(\widehat{N}_t) = 0$:

$$\begin{aligned}E\left(h'(\widehat{N}_t)\widehat{N}_t + \left(h(\widehat{N}_t) - h(0)\right)\right) &= E\left((h'(0) + h''(0)\widehat{N}_t)\widehat{N}_t + \widehat{N}_t h'(0) + \widehat{N}_t^2/2h''(0)\right) \\ &= \frac{3}{2}h''(0)Var\left(\widehat{N}_t\right).\end{aligned}$$

This shows that we need $h''(0) < 0$, i.e. we need that the elasticity is a convex function of employment in this reduced form relationship. Moreover, since $Var\widehat{N}$ is small, $h''(0)$ needs to be large to produce a sizeable wedge. (Recall that in the data, $Var(\xi_t)$ is roughly equal to $Var\widehat{N}_t$.) Given the formula above for h, its second derivative involves second derivatives of

[28] $Cov(g(x), y) = E(g'(x)) \times Cov(x, y)$ if (x, y) is jointly normal.

g. Clearly, our model puts no restriction on this number, and depending on the shape of g it could take either sign and be large or small.

We considered some numerical examples similar to Section 2C, where we pick arbitrarily distributions that match the average employment rate and a given variance of log wages. An example where the wedge works in the right direction can be constructed: pick the second mixture in Table 1 with $r = 10$, change the employment rate to 40%, and assume that employment changes over time according to an AR(1) process (with persistence 0.97 and standard deviation 0.004, but these parameters do not matter much given the formula above). Then the wedge we create is procyclical and its volatility is 10% times the volatility of wages. A natural conjecture is that by changing the shape of the distribution of heterogeneity we can create a wedge that is very volatile and procyclical. However we have not yet found a reasonably calibrated version which generates this result.

We can also look directly in the data and compute the model-implied measure of wedge, i.e. $-\frac{1}{2}\left(\frac{1}{\phi_t} - \frac{1}{\phi^*}\right)\widehat{N}_t$. Next, we compare it to the wedge as usually measured. We find that the model-implied wedge is not volatile enough, with a standard deviation nearly 40 times less than the standard deviation of the data wedge. The correlation between the two was not strong either, and sometimes negative depending on the detrending procedure. (Note that the model-implied wedge is typically larger when ϕ is small on average than when it is large, because it generates larger fluctuations in $1/\phi$.)

To summarize, while we think this topic requires further study, our work suggests that a time-varying elasticity is unlikely to generate a large and procyclical wedge.

C. **CPS Results**

We used monthly Current Population Survey (CPS) data to measure which categories of workers are more cyclical. We considered four possible decompositions of the population:

- Decomposition by gender;

- Decomposition by highest completed schooling (for people 25 years and older);

- Decomposition by age;

- For women, decomposition by marital status, i.e. single, married or separated (divorced or widowed).

These results are available on the following website:

`people.bu.edu/fgourio/laborelastpaper.html`.

D. **PSID Results**

In a previous draft of the paper, we used PSID data to test the prediction that "marginal workers" have a more cyclical labor supply. We derived that implication using an incomplete market model in partial equilibrium. PSID data are annual hours, leading to a time-aggregation problem which limited us in directly testing our hypothesis. Some of the empirical results are however supportive. We found that people with low schooling, low current hours, low family income, or low wages are more procyclical, when one runs a regression of growth rate of hours of each group on the growth rate of aggregate hours. When one runs this equation in levels (i.e. change in average hours on the change in aggregate average hours), the results yield in general a hump-shape. These two sets of results are qualitatively consistent with the results of section 2D.

These results are available on the following website:

`people.bu.edu/fgourio/laborelastpaper.html`.