Precautionary Retirement*

HongSeok Kim†

November 4, 2025

Abstract

With Social Security's trust fund facing depletion, this paper studies how the system's regressive payroll tax shapes labor supply and retirement decision, motivated by an observation that high-income older males work more and longer. We develop a tractable labor supply model with heterogeneous work-costs, demonstrating that the elasticity of labor supply to financial incentives is inherently distributional. High-income workers, further from retirement cutoff, primarily adjust their hours (intensive margin), while those with higher work-costs adjust participation (extensive margin). Our central contribution is the measurement of these income group-specific elasticities, providing a framework to evaluate the distributional effects of policy reforms. We find seniors are 1.71 more elastic than the prime-age workers, implying that raising payroll tax rates may not resolve the solvency of Social Security.

Keywords: Retirement, Labor supply, Work-Costs, Social Security, Payroll Tax.

*

†Iowa State University. Email: hongseok@iastate.edu.

1 Introduction

The long-term solvency of Social Security's Old-Age, Survivors, and Disability Insurnace (OASDI) trust fund is at risk, with projections indicating reserve depletion by 2033 absent legislative intervention. The key contributing factor to this shortage is the erosion of tax base. The share of untaxed earnings has roughly doubled from 10% in 1983 to 17% in 2024. To address this financial challenge, Social Security Administration (SSA) and Congressional Budget Office (CBO) are considering reforms designed to ensure the system's substainability; increasing payroll tax rate or increasing the maximum taxable earnings.

This raises a central question of this paper: Does the payroll tax rate affect the retirement decisions of the seniors? Would increasing payroll tax rate resolves the solvency of the Social Security? The answer is not straightforward. The impact of policy reforms critically depends on how the older workers respond to taxes as their retirement decisions are connected to claiming the benefits. To accurately evaluate the impact of potential reforms, understanding how these fundamentals shape worker behavior across different economic margins is important and ultimately becomes a theoretical and structural question.

The extensive body of literature has thoroughly studied how Social Security benefits influence retirement decisions (Gustman and Steinmeier 1986; Stock and Wise 1990; French 2005; Imrohoroglu and Kitao 2012; Duggan et al. 2023; Yu 2024). The literature includes the analysis of benefit desings: understanding the incentives for early or delaying retirement due to Retirement Earnings Test (RET) or Delaying Retirement Credits (DRC). It also focuses on the institutional details, such as progressive formulas for retirement benefits (Average Index Monthly Earnings or Primary Insurance Amount).

In contrast, how the work-related costs affect retirement decisions remains a relatively underexplored area and has received less attention. Therefore, we shift the focus to the role of work-cost in modeling the labor supply of the seniors. We derive a threshold rule for labor supply covering both intensive and extensive margins. Next, we estimate heterogeneous Frisch ealsticities across income distribution for policy reforms. The main benefits of our approach is as follows: first, it takes account of the extreme work-related costs, such as extreme health or financial shocks. Second, we can decompose the aggregate Frisch elasticities by full-time and part-time, allowing for a larger role of intensive margin. Lastly, our model provides a clear interpretation of the simulation and counterfactuals by fixing the marginal utility of wealth.

We contribute to two strands of literature. First, our paper contributes to the theory of labor supply with intensive and extensive margins by focusing on the seniors. A key feature of retirement is that it is a large, discrete drop in labor income which creates a nonconvexity in individual's budget constraint. To make a retirement an endogenous choice, we need either an assumption or friction. We develop a tractable labor supply model with non-pecuniary work-costs which are drawn from Extreme Value distributions. This modeling approach is built upon growing evidences on the importance of non-monetary aspects of jobs, such as disutility of work (Kaplan and Schulhofer-Wohl 2018), non-market work (Aguiar and Hurst 2016), leisure (Aguiar et al. 2021), and mental health (Abramson et al. 2024).

Second, we contribute to the long-standing debate on reconciling micro- and macro-aggregate Frisch elasticity of labor supply (Hansen 1985; Rogerson 1988; Prescott 2004; Chang and Kim 2006; Chetty et al. 2011, 2013; Ljungqvist and Sargent 2011; Erosa et al. 2016; Kleven et al. 2025). The struggle comes from a difference between short-term, cross-sectional views and long-term, life-cycle perspectives. Kleven et al. (2025) argues that quasi-experimental variations cannot capture the long-term dynamic returns to an effort. The consensus in the literature is that there exists no single aggregate Frisch elasticity, making it important which type of heterogeneity to capture. Our paper studies an ideal setting where cross-sectional variations and dynamic variations coincide. We focus on the seniors whose human capital accumulations are largely completed so that their dynamic returns to an effort are small. Under this setting, we can estimate heterogenous Frisch elasticities without worrying about the time effects.

The central argument is that the aggregate Frisch elasticity of labor supply - in response to the financial incentives for working - varies significantly with income. The distributional effect arises because the retirement decision is shaped by the Social Security's payroll tax rate and the costs of continuing employment among the seniors. We estimate income group-specific elasticities which can be effectively used for evaluating policy reforms.

We first document a key empirical evidence on the labor supply of seniors using Health and Retirement Study (HRS) data. High-income seniors work more, and for longer hours, than their low-income counterparts, even after claiming the Social Security benefits. Such difference in labor force participation rate and average working hours is particularly stark after age 70. We focus on the trend after age 70 because all individuals are expected to have received the retirement benefits. This allows us to isolate the effect of Social Security benefits on the labor supply of seniors. The trend is robust even when we control for wealth and birth cohort. Therefore, the income heterogeneity across seniors will be at the core of the modeling assumption.

Second, we develop a tractable labor supply model incorporating both intensive and extensive margins. The key innovation of the model is its focus on the non-monetary or non-pecuniary costs of working which can be effectively modeled using Extreme Value distributions. For example, we can think of large out-of-pocket medical expenses due to health shocks, unexpected expenditures for relocations for job transition, or large fixed costs of re-entering the labor market after unemployment spell, all of which affect the (dis)utility of working. Such limiting distributions are well-suited to capture the maximum effect of various shocks. This approach provides a microfoundation for labor decisions grounded in the nonpecuniary aspects of work.

There are two primitives that discipline the model. First, there is a fixed time cost that must be paid overhead to participate (commuting time, coordination between workers, or set-up costs) which creates a nonlinear mapping from hours to labor (Rogerson and Wallenius 2009, 2013). This creates a nonconvexity that makes "retire" and "work" as options. Second, each household face a multiplicative work-cost (or disutility) shock z from an Extreme Value distribution. The marginal return to an additional hour or work acts as a sufficient statistic through which policy and prices play a role. Given the marginal return to an hour, an individual choice depends on a threshold: an individual works if the shock z lies below a cutoff $\bar{z}$, and retire otherwise. We can then aggregate individual labor supply over the distribution.

Aggregating across households makes the mapping from policy to observables transparent. A reform that raises (lowers) the net return of an hour shifts the entire window of full-time and retirement threshold to the right (left), increasing (decreasing) participation and raising (lowering) the share of workers at the full-time region and average working hours. The heavy right tail of extreme value distribution is economically important. It implies that even in the high-income older workers, there remains a meaningful set of individuals whose hours can adjust when incentives change, while groups closer to non-participation respond more on the extensive margin. In this way, a single net-return measure maps policy changes into distribution-specific predictions for participation, the incidence of full-time work, and average working hours.

Third, we calibrate the model to the cross-sectional moments of the U.S. in 2017 and 2018. We target the moments of individuals age 65 and older across income quintile. We match the shape of earnings per worker and full-time share for each income quintile as well as aggregate full-time share of seniors. We find a parameter value that governs the curvature of labor supply somewhat reasonable compared to the literature. The annualized fixed time cost is also comparable to that of French (2005).

We analyze a steday-state economy. Given the calibrated parameters, we simulate the labor supply response to a small perturbation in the marginal return to an hour. We then use the finite difference method to compute the Frisch elasticity for each income group. In this model, participation is endogenous, allowing for both an intensive and extensive margin response. We use the structural relation to define the participation threshold, which would be updated accordingly. The changes in aggregate hours is the sum of change in conditional hours worked weighted by employment rate, and the change in employment rate weighted by the average hours of those employed. Therefore, the aggregate Frisch elasticity for the economy is the employment-weighted average of these group-specific total elasticities.

Lastly, we conduct two counterfactuals to understand the behavior of seniors. For the first counterfactual, we eliminate the payroll tax rate to study how seniors adjust their labor supply. Since we do not have firms in the model, we consider individuals face a flat payroll tax rate 12.4%[1]. When we eliminate the payroll tax rate to zero, we observe 3.6% increase in total working hours and 3.9% increase in share of full-time employment. Once we decompose aggregate changes of working hours by income

1. For the highest income distribution, not all individuals pay payroll taxes because their earnings are above the earnings cap. The maximum taxable earnings in 2024 is $176,100 and the cutoff for the higest income distribution in our model is $110,000. However, the share of individuals whose earnings above the cap is around 9%. Given the smaller share of seniors whose earnings are above the cap, we assume every seniors pay the payroll tax.

quintile, we see the largest increase (11.9%) from the lowest income seniors and the smallest increase (1.9%) from the highest income seniors. The same applies for the full-time share, 14.5% and 2.1%, respectively. For the extensive margin, the fraction of retirees adjusting to enter the labor force is about 35% from the lowest income group and 7% from the highest income group. In our economy, we observe 14% of aggregate increase in the labor force.

For the second counterfactual, we recalibrate the model to match the cross-sectional moments of the prime-age. We compare changes of aggregate labor supply of seniors and prime-age workers. The primary objective of this exercise is to understand whether seniors respond to taxes than the prime-age. We observe 2.3% increase in aggregate labor supply when payroll tax rate is eliminated. Given that we observe 3.6% increase for the seniors, we infer that the seniors are 1.71 more elastic than the prime-age.

The results have policy implications for potential reforms of Social Security. The uniform payroll tax can be regressive on seniors. Not only the presence of maximum taxable earnings make the system regressive, but also the low income seniors are disproportionately affected regarding labor decisions. Hence, increasing payroll tax rate can incentive the low and middle-income seniors to exit the labor force and claim Social Security benefits early. This outcome, with fewer workers paying in and more retirees drawing out, would worsen, not improve the pressure on Social Security's finances.

The rest of the paper is organized as follows: Section 2 provides empirical evidences on the labor supply of seniors. Section 3 develops the labor supply model. Section 4 calibrates the model and results. Section 5 studies counterfactuals of the model. Finally, Section 6 concludes.

2 Labor Supply of the Seniors

2.1 Defining Retirement

Before we discuss the data, we ask the following: how should we define retirement? The timing of retirement is not clear-cut because there exists no single definition. Social Security Administration (SSA) considers *claiming the benefits* as retirement, while the legal dictionary defines retirement as *full withdrawal from the labor force*.

We argue that none of these definition fully capture the retirement behavior of seniors. We find that 23% of seniors who already claimed the retirement benefits still remain in the labor force, using the Health and Retirement Study data from 1992 to 2010. This is substantiated by the findings of Shoven et al. (2017) which documents that one-thirds of respondents in their survey kept working upon claiming the benefits.

2.2 Data

We use the Health and Retirement Survey (HRS) from survey years from 1992 to 2010 provided by RAND. HRS is a biannual large scale longitudinal panel study that surveys a natinoally representative sample over age 50 in the U.S. It provides information on employment status and retirement decisions of indviduals, making this a suitable data to study our question, particularly for the seniors.

We construct our sample from the HRS data to isolate the labor supply decisions of seniors. The sample consists of male household heads aged 62 and older. We start at age 62 because it is the eligible age for early retirement. We further restrict the sample to individuals who have already claimed benefits but continue to work (reporting positive hours and a "working" status). This allows us to analyze labor decisions after the claiming decision has been made. Lastly, we stratify this sample by income percentiles in each wave using weekly wages.

Throughout this paper, I refer individuals aged 70 and older who claim Social Security benefits and continue to work as *precautionary retirees*. To distinguish them from the broader elderly population (all those 65 and older), I use the term "seniors" as a shorthand for these precautionary retirees. This approach is informative when considering the full age range of older workers. Once workers reach age 70, for example, there is no longer a financial incentive to daly claiming Social Security benefits, because Delayed Retirement Credits (DRCs) no longer apply. Therefore, individuals receive benefits regardless of whether they decide to continue working or leave the workforce. This provides a clear threshold to investigate the data, as claiming-related rules are no longer a main driving force inducing the seniors to remain in the labor market after age 70.

(a) Cohort: 1920 - 1939 **(b)** Cohort: 1940 - 1959

FIGURE 1. Extensive Margin of Labor Supply by Earnings

Notes. The rank is decomposed into five earnings quintiles. The data sample consists of male household heads who defined themselves as working, conditional on receiving Social Security benefits. Data source: Health and Retirement Survey (HRS), 1992-2010.

2.3 Intensive & Extensive Margin of Labor Supply

Figure 1 and Figure 2 show the extensive and intensive margins of seniors' labor supply. The analysis focuses on the trend after age 70, as retirement timing before this age is heavily influenced by Social Security rules and is subject to selection bias. After age 70, a clear trend emerges: high-income seniors work more and for longer years than their low-income counterparts. This pattern is consistent across birth cohorts, while the labor supply gap between the high- and low-income is more pronounced in the older cohort.

Extensive Margins of Labor Supply Figure 1 shows the share of individuals age 62 and older who work positive annual hours and report a working status. Using the weekly wages, we stratify the sample into five income quintiles. We clearly observe that there is a distinct labor behavior between the high- and low-income quintile. The trend generally peaks around age 70 and gradually declines. In the lowest income group (Quintile 1), the share declines starting from age 62. For the highest income quintile, employment shares of the senior Social Security beneficiaries remain more or less constant after age 70. Figure 1 compares two birth cohorts: individuals born in 1921–1939 and those born in 1940–1959. We track each cohorts labor force participation rates over time by income quintile. The a similar trend holds for different birth cohorts.

(a) Cohort: 1920 - 1939

(b) Cohort: 1940 - 1959

FIGURE 2. Intensive Margin of Labor Supply by Earnings

Notes. The rank is decomposed into five earnings quintiles. The data sample consists of male household heads who defined themselves as working, conditional on receiving Social Security benefits. Data source: Health and Retirement Survey (HRS), 1992-2010.

Intensive Margins of Labor Supply Figure 2 presents annual working hours among the Social Security beneficiaries aged 70 and older, differentiated by income quintiles and birth cohorts (1920-1939 and 1940-1953). A strikingly similar pattern emerges between the low- and high-income groups. While the intensive margins of labor supply among the low-income senior steadily declines, those of the high-income seniors remain far above. When we compare the intensive margins across different birth cohort, even though the gap is narrowed, the trend is robust.

One clear inference we make from these empirical facts is that the timing of retirement is heterogeneous across income distribution of the senior workers. We observe that high-income seniors are more likely to remain in the labor force even after claiming Social Security benefits. This could be due to better health outcomes, type of occupations they work, or preferences for working, the answer which this paper does not address. Instead, we consider the income heterogeneity of seniors seriously which will be at the core of the modeling assumption in Section 3.

2.4 What are Different About the Labor Supply of Seniors?

Part-Time Jobs Unlike prime-age workers, we highlight the role of part-time jobs for those who are eligible for retirement. There are increasing evidences of "bridge jobs"

TABLE 1. Distribution of Male Annual Hours ($\leq$ 1750) by Age (PSID)

Age	60	61	62	63	64	65	66	67	68	69	70
Share of individuals	19%	22%	22%	23%	26%	26%	33%	28%	25%	21%	20%

Notes. These represent the share of individuals of male households who work less than 1,750 but show positive working hours. All the numbers are calculated from the Table 1 of Rogerson and Wallenius (2013) by summing up the share between (0, 1750). CPS considers 35 hours and more as a criteria for full-time job.

or part-time jobs before completely exiting the labor force (Maestas 2010; Shoven et al. 2017; Coile 2025). Compared to a robust empirical finding of concentrated 40-hours per week of prime-age workers (Bick et al. 2022), part-time jobs should not be disregarded and be considered as an important factor to understand retirement behaviors of the seniors.

Table 1 shows the distribution of annual working hours by male households between age 60 and 70. Given that CPS considers working 35 hours and more per week as a full-time job, we can understand the numbers in Table 1 as a share of seniors who work part-time jobs. The share of seniors for each age who are working part-time is between 20% and 28%. Rogerson and Wallenius (2013) denotes that 75% of seniors are working either full-time or zero hours, focusing the analysis on the extensive margins of labor supply.

However, combined with the rising longevity and demands for working longer among the seniors (Ameriks et al. 2020), the share of working part-time is likely to increase. Part-time jobs are becoming prevalent in the labor market of seniors since firms are more flexible and have more discretions in making contracts. As of 2024, about 38.3% of employed Americans age 65 and older worked part-time. The rate is much higher than among younger age groups: 14.2% for ages 55-64 and 11.1% for age 25-54.

Extreme Shocks Another key distinction between seniors and prime-age workers is the higher prevalence and severity of health shocks. This has made the analysis of these shocks, particularly their causal effect on the retirement decision, an indispensable area of research in retirement literature.

We document a rising out-of-pocket (OOP) medical expenses as an evidence for the importance of health shocks particularly for the seniors. Table 2 shows the median annual long-term care costs. Long-term care is considered a primary out-of-pocket medical expenses for the US seniors. Comparing the costs from 2010 and 2024, the costs of care types all increased significantly, except for nursing home services.

Table 2. Median Annual Long Term Care Costs (2024 dollars)

Care Type	**2010**	**2019**	**2024**	**Δ (2010 - 2024)**
Homemaker Services	\$59,246	\$63,165	\$75,504	+27.4%
Home Health Aide	\$62,538	\$58,694	\$77,792	+24.4%
Assisted Living Facility	\$56,847	\$59,646	\$70,800	+24.5%
Nursing Home (Semi-Private)	\$108,166	\$110,620	\$111,325	+2.9%
Nursing Home (Private)	\$120,243	\$125,403	\$127,750	+6.2%
Adult Day Health Care	\$22,442	\$23,926	\$26,000	+15.9%

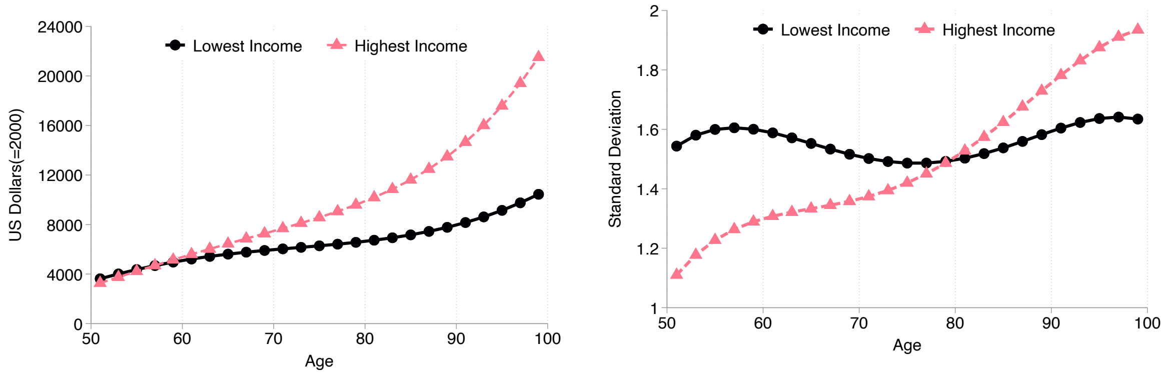
Notes. All amounts expressed in constant 2024 dollars using the CPI (BLS). Real change shows cumulative inflation adjusted growth from 2010 to 2024. Data Sources: Genworth and CareScout Cost of Care Surveys (2010, 2019, 2024)

We also use Health and Retirement Study (HRS) data to estimate the expected out-of-pocket expenses and corresponding risks. To do so, I follow a regression model from Nakajima and Telyukova (2025):

$$\log(\texttt{OOP}_{it}) = \beta_0 + \beta_1 \texttt{Age}_{it} + \gamma X + \varepsilon_{it}$$

where $\texttt{OOP}_{it}$ denotes out-of-pocket expenses for individual i at time t, X is a vector of control variables, and health status is held constant at ‘good’.To capture the significant tail risk in medical spending, a log-normal distribution for out-of-pocket expenses is assumed.

Figure 3 shows estimate results of medical risks by showing how expected out-of-pocket medical expenses and their associated risks change by age. One notable finding is that the expected out-of-pocket medical expenses and associated risks are increasing in age for the wealthy seniors, while those are more or less constant for the poor seniors. Therefore, we can deduce that health-related costs have large, heterogeneous impacts on the effective labor of the seniors.



(a) Expected OOP Medical Expenses

(b) Log Standard Deviation of OOP Medical Expenses

FIGURE 3. Estimation of Expected Out of Pocket Medical Expenses and Risks

Notes. This figure replicates a figure of Nakajima and Telyukova (2025).The graphs are conditional on the subjective health status "good" from HRS data.

3 A Labor Supply Model of Seniors

3.1 Environment

In this section, we develop a tractable labor supply model of seniors with heterogeneous work-related costs. We analyze a social planner's problem within a partial equilibrium framework, taking wages and interest rate as given. The model assumes complete markets, which provides full insurance. Since our focus is on seniors, we model their life-cycle from age 65 to 100, normalizing this time horizon to the interval [0,1].

Preferences We build upon the time-average model of Ljungqvist and Sargent (2014). To make a retirement an endogenous choice, we introduce a fixed time cost which generates a nonconvexity in the budget constraint (Rogerson and Wallenius 2009, 2013). Each individuals choose consumption $c_t(z)$ and working hours $h_t(z)$ to maximize their present-value expected utility at each period given the realization of shock z

$$\int_0^1 e^{-\rho t}\left[\int \Lambda(z)\Big(u(c_t(z)) - z\cdot\phi\big(h_t(z)+\underline{h}\big)\Big)dF_Z(z)\right]dt \tag{1}$$

where utility $u(\cdot)$ is an increasing, concave function and working costs $\phi(\cdot)$ is increasing, convex function. Individuals have subjective discount factor ρ. Taste shocks or disutility of working z is drawn from a time-invariant Extreme Value distribution F_Z with density f_Z. There is a fixed time cost $\underline{h}$ which must be paid overhead. We are thinking of commuting costs, coordination among workers, and health status. This nonlinear mapping from hours to labor generates "working" and "retirement" as choice variables.

The social planner assigns a Pareto weight $\Lambda(z)$ on each type z which

$$\int \Lambda(z) dF_Z(z) = 1$$

holds on average. Assuming a different distribution on $\Lambda(z)$ would introduce ex-ante heterogeneity across seniors. To simplify the exposition, however, we consider $\Lambda(z)$ is constant and equal to one across all types.

Resource Constraint The planner maximizes an objective function (1) subject to a present-value resource constraint,

$$\int_0^1 e^{-rt} \left[\int \left((1-\tau) w_t h_t(z) - c_t(z) + T_t \right) dF_Z(z) \right] dt \geq 0. \tag{2}$$

with wage rate w_t per unit of labor hour, interest rate r, and a flat payroll tax rate τ.

Government All the proceedings of taxes are rebated as lump-sum transfers (Prescott 2004),

$$T_t = \tau \int w_t h_t(z) dF_Z(z).$$

3.2 Equilibrium

The social planner maximizes the present-value expected utility (1) subject to a present-value resource constraint (2). Let μ be the Lagrangian multiplier for the resource constraint. We assume that the subjective discount rate ρ is equal to the interest rate r. This "time-averaging" assumption allows us to disregard the effect of timing of decisions. For instance, if

The optimality condition for consumption is read as

$$\Lambda(z)u'(c_t(z)) = \mu e^{-(r-\rho)t} = \mu \tag{3}$$

which equates the marginal cost and benefit of consumption. Clearly, the right-hand side is independent of shock z and time t. By the assumption $\Lambda(z) = 1$, the consumption is constant, $c_t(z) = c^*$, over a pair (t, z). This full-insurance result comes from the complete market structure: the planner would insure the consumption across different individuals regardless of their extreme value taste shocks at each period.

The optimality condition for labor hours is read as

$$\Lambda(z)z\phi'(h_t(z) + \underline{h}) = \mu e^{-(r-\rho)t}(1-\tau)w_t = \mu(1-\tau)w_t \tag{4}$$

which equates the marginal cost and benefit of an additional working hour. By the time-averaging assumption and constant consumption, the marginal return to an hour is also constant at each period,

$$(1-\tau)w_t u'(c^*) = \Psi_t$$

which we denote it as Ψ_t given the wage rate w_t.

We assume an isoelastic labor supply function

$$\phi(h_t(z)) = \frac{1}{1+\frac{1}{\eta}}\left(h_t(z)\right)^{1+\frac{1}{\eta}}, \quad \eta > 0$$

with a parameter η that governs the shape of labor supply. Using this functional form and rewriting (3) and (4), we derive an intensive margin of part-time labor supply

$$h^{\text{int}}(z) = \left(\frac{\Psi_t}{z}\right)^{\eta} - \underline{h}, \quad \Psi_t = (1-\tau)w_t u'(c^*). \tag{5}$$

Note that working hours is a decreasing function in shock z and the range is $[0, 1-\underline{h}]$. For sufficiently low disutility of working, $z < \underline{z}(t)$, individuals will work full-time $1-\underline{h}$ where $\underline{z}(t) = \Psi_t$ is the cutoff between full-time and part-time. This cutoff is derived from the optimality condition for an interior optimum of labor supply. Recall $\Psi_t = z\phi'(x)$ for

$x = h + \underline{h} \in [0, 1]$. The full-time constraint binds when the interior solution is $x \geq 1$. Setting $x = 1$ in the first-order condition and solving for z gives

$$\underline{z} = \frac{\Psi_t}{\phi'(1)} = \Psi_t$$

under an isoelastic labor supply function.

Likewise, we can derive a threshold rule that determines the extensive margin between working and retirement. The threshold for working and retirement is

$$\overline{z}(t) = \frac{\Psi_t}{\left((1+\eta)\underline{h}\right)^{\frac{1}{\eta}}} \tag{6}$$

such that

$$h_t(z) = \begin{cases} \min[1 - \underline{h},\, h^{\text{int}}(z)], & \text{if } \; z < \overline{z} \\ 0, & \text{if } \; z \geq \overline{z}. \end{cases}$$

To summarize, we derive a threshold rule that maps the shock distribution to labor supply

$$h_t(z) = \begin{cases} 1 - \underline{h}, & \text{if } z < \underline{z}(t) \\ \left(\frac{\Psi_t}{z}\right)^{\eta} - \underline{h}, & \text{if } z \in [\underline{z}(t)\,,\, \overline{z}(t)] \\ 0, & \text{if } z > \overline{z}(t). \end{cases} \tag{7}$$

3.3 Aggregate Labor Supply

Given the threshold rule (7), we can compute the aggregate labor supply for each period by integrating the share of full-time and part-time,

$$\begin{aligned} H_t &= \int h_t(z) dF_Z(z) \\ &= \int_{z \leq \overline{z}} \min\left[1 - \underline{h},\, \left(\frac{\Psi_t}{z}\right)^{\eta} - \underline{h}\right] dF_Z(z) \\ &= F(\underline{z}(t)) + \Psi_t^{\eta} \int_{\underline{z}(t)}^{\overline{z}(t)} z^{-\eta} dF(z) - \underline{h} F(\overline{z}(t)). \end{aligned} \tag{8}$$

The first term denotes the share of individuals in the full-time region. The second term denotes the share of individuals in the part-time region. The third term denotes subtraction of fixed time costs of all individuals in the working region. Figure 4 shows the mapping graphically.

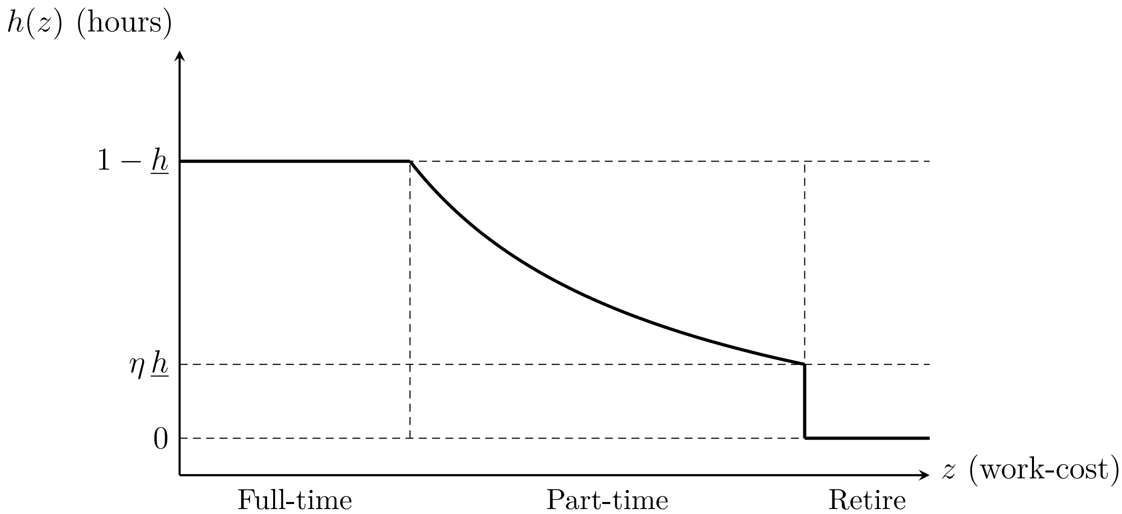


FIGURE 4. Mapping from Shock to Labor Supply

Now we can derive aggregate Frisch elasticity of labor supply using the threshold rule (8).

$$\varepsilon_{\text{total}} = \frac{d\ln(H_t)}{d\ln(A_t)} = \underbrace{\frac{1}{H}\int_{A_t}^{\overline{z}(t)} \varepsilon_{\text{indiv}}(z)h(z)dF_Z(z)}_{\text{Intensive margin}} + \underbrace{\frac{\eta\underline{h}}{H_t}\,\overline{z}(t)f_Z(\overline{z}(t))}_{\text{Extensive margin}} \tag{9}$$

with an individual Frisch elasticity

$$\varepsilon_{\text{indiv}}(z) = \eta\left(\frac{h(z)+\underline{h}}{h(z)}\right)$$

All derivations are in Appendix (B.1). This implies that intensive Frisch elasticity is endogenous to working hours. Households working longer hours are less inelastic than those working shorter hours, the same finding as Karabarbounis (2016). However, this model provides another angle from which we interpret the Frisch elasticity. When fixed time cost is operative, the optimal working hours h will depend on $\underline{h}$. The higher the fixed costs, higher the Frisch elasticity. Also, with the fixed time cost $\underline{h}$, the optimal working hours h^{int} is bounded below by $\eta\underline{h}$ and the maximum Frisch elasticity among workers is finite. Hence, an intensive Frisch elasticity of individual labor supply is between 0 and $1+\eta$. A further explanation for the lower bound which is derived from cutoff rule will be discussed below.

Remark 1. *Suppose the fixed time cost of labor supply is zero, $\underline{h}=0$. Then the extensive margin becomes zero and full-time region no longer exists. Then every households work $h(z)=(A/z)^{\eta}$ hours. In this case, the total Frisch elasticity becomes*

$$\frac{dH}{d\ln(A_t)} = \int \eta h dF = \eta H$$

or

$$\varepsilon_{total} = \frac{d\ln(H)}{d\ln(A_t)} = \eta = \varepsilon_{indiv}$$

where micro and macro elasticities coincide. This was the starting point on the literature of reconciling micro- and macro-aggregate Frisch elasticity. Were it not for the fixed time cost, the structural parameter is linked to the aggregate Frisch elasticity. Our model can explain both small individual Frisch elasticity and large aggregate Frisch due to extensive margin.

Once the fixed time costs are present, the total elasticity of labor supply is larger,

$$\eta\left(\frac{h+\underline{h}}{h}\right) > \eta$$

because the entry term is added. Note that if many households are binding at the full-tiem working hours constraint, the intensive margin is smaller and the aggregate Frisch elasticity can be smaller.

3.4 Understanding the difference

One major concern is that why do we need a labor supply model of seniors. We first investigate a standard macro framework to study the labor supply of prime-age workers as shown in Figure 5. This framework is based on a robust empirical fact that more than 50% of employed prime-age workers is heavily concentrated in 40-hours per week

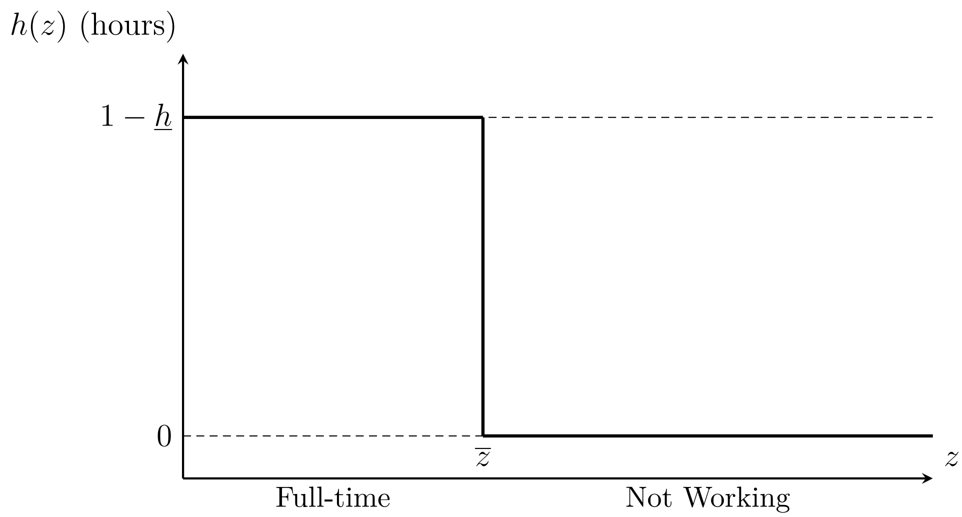


Figure 5. A Standard Framework for Prime-age Workers

(Bick et al. 2022). Therefore, many quantitative general equilibrium macro models assume a strict retirement rule in which all individuals retire at age 65. For instance, if an individual retires, then all changes of aggregate labor supply will be interpreted as extensive margins.

However, the labor supply of seniors are not as the same as the prime-age. As emphasized in previous section, the role of part-time jobs are increasingly pervasive in seniors' labor supply decision, let alone retirement. Our model is able to capture this gradual transition to retirement which is displayed in Figure 4. In this case, not all changes of labor supply are extensive margin, but interpreted as an intensive margin.

3.5 Alternative Mechanisms

We use the model framework above to interpret why the high-income seniors are working longer and more than the poorer counterparts. For comparability, we investigate the role of Frisch elasticity where substitution effect is only considered.

Mechanism Introducing heavy-tailed Extreme Value (EV) distributions shape the mapping from marginal return to an hours Ψ to labor supply in two ways. One mechanism is that they affect full-time share among workers. The probability of full-time is

$$\text{FT} = \frac{F_Z(A)}{F_Z(\overline{z})} = \frac{F_Z(\Lambda\overline{z})}{F_Z(\overline{z})}, \quad \text{where } \Lambda = ((1+\eta)\underline{h})^{1/\eta} \in (0,1).$$

For a Fréchet distribution $F(z) = \exp(-(1/z)^{\alpha})$ where α is a shape parameter, FT rises as $\overline{z}$ moves right. That means higher Ψ pushes more share of workers to a full-time region $z \leq \Psi$. Another mechanism is that EV distributions affect the aggregate elasticity of labor supply via (9). With heavy tails, there remains a non-negligible probability at high $\overline{z}$, so that intensive term in $[\underline{z}, \overline{z}]$ is sizeable.

Effective Shadow Price of Hours This case represents in which $\Psi_{j,t}$ is higher for both income states $j \in \{\text{High, Low}\}$. Since we fix $u'(c_t)$ for all income states, higher wages increases A_t which, in turn, increases intensive working hours and participation rates for high-wage and low-wage workers.

Since $w_H > w_L$ by assumption, we have $\Psi_{H,t} > \Psi_{L,t}$ and as a result, the changes of intensive working hours between workers

$$h_{H,t} - h_{L,t} = (\Psi_{H,t} - \Psi_{L,t})z^{-\eta} > 0$$

is larger for the high-income workers for given taste shock z. Intuitively, the substitution effect for high-income worker is larger as the opportunity cost of leisure is higher.

Similarly, the high-income workers participate more in the labor market than the poorer counterparts

$$P_{H,t} - P_{L,t} = F_Z(\overline{z}_{H,t}) - F_Z(\overline{z}_{L,t}) > 0$$

by the equation (6).

Fixed Time Costs We can argue that fixed time costs can vary by income level: $\underline{h}_H < \underline{h}_L$ or vice versa. We consider the case $\underline{h}_H < \underline{h}_L$ holds which is an innocuous assumption if fixed time costs - the rising trend of Working From Home (WFH) reduces a commuting costs for the high-income professionals - or it could be the flexibility of schedule a job can provide.

In this case, an increase of additional fixed time costs would lead to higher participation rate for the high-income workers,

$$\frac{\partial H_t}{\partial \underline{h}_H} = -\frac{\overline{z}}{\eta \underline{h}_H} > -\frac{\overline{z}}{\eta \underline{h}_L} = \frac{\partial H_t}{\partial \underline{h}_L}$$

while the intensive response of labor supply is equal for both types of workers.

Distribution of Taste Shocks Suppose the distribution of taste shocks z are heterogenous for different type of workers. For instance, consider the average quality of job amenities provided to high-income workers is significantly better than the low-income jobs. Then, the average value of z is lower or the density is thinner near the tail.

Assume that F_Z^L first-order stochastic dominates F_Z^H. That is $F_Z^L(z) \geq F_Z^H(z)$ for all z. To fix an idea, assume that both types earn equal wages (holding A_t constant), but only the distribution of z differs. Then, for any given A_t and $\underline{h}$, an extensive response of labor supply is higher for the high-income type, or

$$P_L = F_Z^L(\overline{z}) \leq F_Z^H(\overline{z}) = P_H$$

at the cutoff $\overline{z}$. Similar argument applies to the intensive response of labor supply as we integrate (7) on the same support over different distributions.

Taking Stock The objective is to build a theory that can explain why high-income workers work more and for longer hours (intensive and extensive margin) than the poorer counterparts. As noted by Ferraro and Valaitis (2024), a standard equilibrium model cannot explain such facts. That is, the wealth effects dominate the endogenous labor supply decisions and leisure as a normal good, making wealthier households to consume more and work less.

The model can handle both intensive and extensive margin of labor supply and derive the total Frisch elasticities. The core idea is similar to Rogerson and Wallenius (2009) where the planner can sort workers by the level of disutility, creating a threshold rule. Given such property, this model can succesfully explain why the high-income individuals work more and longer than the low-income individuals.

4 Quantitative Analysis

In this section, we estimate the model parameters of the labor supply with Extreme Value distributions. The model is calibrated to the US economy in 2017 before COVID-19. We take two steps for estimation. First, I estimate the necessary parameters without the presence of payroll tax. We find that the model captures an increasing profile of earnings index (intensive margin) and the rising fraction of participation (extensive margin) across income distributions. Next, given the model parameters, we introduce a regressive payroll tax in the model and compare the results.

4.1 Data

We primarily rely on two data sources for calibration. First, we use the data from Congressional Research Service (2022) (henceforth, 'CRS'). CRS links Health and Retirement Study (HRS) data to an administrative data from Social Security administration (SSA). Therefore, the objective in this section is to estimate parameters that can replicate the empirical regularities documented by CRS.

In order to avoid the effect of COVID-19, we use 2018 HRS data which contains labor market information of the elderly on 2017. CRS provides cross-sectional target moments of age 65 and older by household income quintiles. Compared to prime-age workers, labor supply of the seniors are scarce. Although HRS provides a detailed, high-quality data, the sample size gets significantly lower when it is conditioned on several dimensions. Therefore, we rely on the estimates provided by the CRS given its larger sample sizes and linkage to SSA administrative data.

We use the share of individuals, age 65 and older, with household earnings as a proxy for "worked during year" by income quintile.[2] In other words, this assumes that any individual with a positive earning would have participated in the labor market. Table 3 documents a positive earnings-income gradient. The higher the income distribution, the more the share of the seniors who have participated in the labor market.

Second, we use the data from Bureau of Labor Statistics (BLS) on the share of

2. CRS documents the quintile cutoffs for age 65 and older: $\$21,840$, $\$39,700$, $\$64,295$, $\$110,897$ as an upper bound for Q1-Q4.

share of age 65 and older who are working full-time in 2018[3]. We use the following formula

$$\text{Share of Full-Time Workers (65+)} = \frac{\text{Employed 65+ who work 35+ hours}}{\text{All employed 65+}}$$

where working more than 35 hours is counted as full-time in CPS. The share of employed individuals age 65 and older working full-time is 61%. This aggregate share is used to discipline the population-weighted intensive margin of the labor supply of all the seniors.

4.2 Calibration

We assume that individuals with earnings are the ones who have participated in the labor market. Then, we take the observed participation rate (the first row of Table 3) as an extensive margin target p_q for each income quintile q. That is

$$p_q = \Pr[\text{Earnings} > 0]$$

for $q \in \{1, 2, 3, 4, 5\}$. For example, p_1 is 0.095 and p_5 is 0.672. We use these point estimates for extensive margin of labor supply for each income quintile. Given p_q and a common probability distribution F_Z, we invert to retrieve a group-specific values

$$A_q = ((1+\eta)\underline{h})^{\frac{1}{\eta}} F_Z^{-1}(p_q)$$

that automatically matches the labor force participation rate p_q by construction.

To discipline the intensive margin of labor supply for each income quintile, we use two target moments. First, we use a term "earnings intensity" by dividing the second row to the first row of Table 3. This term

$$\text{Earnings Intensity} = \frac{\text{Earnings Share Per Person}}{p_q} \propto \text{Earnings Share Per Worker}$$

is now proportional to *earnings per worker* within each quintile up to a constant scale.

3. https://www.bls.gov/blog/2019/how-are-our-older-workers-doing.htm

Table 3. Target Moments by Income Quintile

Income Quintile	**Q1**	**Q2**	**Q3**	**Q4**	**Q5**
Share of individuals with earnings	9.5%	21.5%	37.2%	47.4%	67.2%
Share of earnings to total income	2.9%	7.0%	13.6%	19.6%	32.6%

Notes. All income distribution is divided into five quintile. Q1 denotes the lowest 20% and Q5 the highest 20% of income distribution. First row denotes the share of individuals (65 and older) who receive any earnings in 2017. Second row represents the share of earnings to the aggregate household income. Source: Figure 5 and Figure 6 in Congressional Research Service (2022)

Only after normalizing the average value within each quintile to 1 can we compare across income quintiles while the common scale term washes out. This way allows us to estimate the shape parameter of the Extreme Value distribution F_Z.

Since we do not have data on individual wages and working hours by quintile, a reformulation is necessary because we do not have data on wages and consumptions for each individuals. We cannot match the level of marginal utility of hours, or $A = w \cdot u'(c)$. However, earnings intensity enables us to match the shape or pattern across quintiles, thereby producing a testable implication of the model.

4.3 Results

We present the role of heterogeneity in distribution to produce responses across income distributions. Figure A.2 shows that the intensity response is constant across income quintile absent heterogeneity.

Then we estimate parameters via method of moments. The key parameters to be estimated are $\{\eta, \underline{h}, \alpha, \beta\}$ where α is a shape parameter estimated for Fréchet distribution and β for log-Gumbel distribution. Each households draw a taste shock for work z at the beginning of the period from an Extreme Value distribution. For a given income quintile q, individuals work if their draw is below a quintile-specific threshold $\overline{z}_q$, and are non-participants otherwise. Conditional on working, they choose hours according to (7).

Table 4 shows the estimated parameters $\{\eta, \underline{h}, \alpha, \beta\}$ for both distributions. The shape parameters are important to a degree in which it magnifies how the differences in participation rate map into differences in the cutoff $\overline{z}$ and hence to the net return.

For Fréchet distribution, $\alpha = 1.2828$ implies a small heavy right tail, meaning more probability of very high taste shocks z. For log-Gumbel distribution, it is a re-scaling of Fréchet but with an effective shape parameter $\beta = 3.884$. If we assume $\alpha = 1/\beta$, two distributions are estimating the same distribution.

Two important parameters are η which governs the shape of intensive working hours, and $\underline{h}$ which makes labor participation costly. For the labor supply curvature, the larger the η, the stronger the hours responses. η determines the magnitude of Frisch elasticity with the fixed cost of work $\underline{h}$.

The estimated fixed cost of work ($\underline{h}$) from our model is 0.3999. To translate this parameter into an intutitive metric, we convert it into hours per week based on a standard 40-hour full-time schedule documented by Bick et al. (2022). Using the formula,

$$\text{Fixed hours/week} = \frac{\underline{h}}{1-\underline{h}} \times (40 \times 0.6 + 20 \times 0.4)$$

where 60% of seniors work full-time and 40% of them work part-time. I assume part-time working hours is the half of full-time. We find that the fixed cost of work is equivalent to 21.3 hours per week for Fréchet, and one hour per week for a log-Gumbel distribution. We annualize the fixed time cost by multiplying 40 weeks[4] for Fréchet which gives 852. This is similar to 885 hours calculated by French (2005). Hence, we use Fréchet distribution afterwards.

We highlight that because all individuals at each income group face the same fixed time costs, the share of fixed costs for each group is heterogeneous. For instance, the high-income old-age households work more hours on average, therefore fixed time cost amounts to a smaller fraction of their total hours. This decreasing trend is what makes the labor supply responses so different across income group.

Figure 6 shows the model performance. We provide two results: the shape of earnings intensity which acts as a proxy for intensive margin response, and the share of full-time workers of each income group. Since we use the ratio of different income groups (Q4/Q2 and Q5/Q4) for calibrating the parameter, it automatically matches the data.

4. National Institutes of Health documents that a majority of seniors work fewer weeks than full-year schedules for younger adults, and Bureau of Labor Statistics also reports that part-time working is prevalent for seniors than any other groups. Therefore, we assume 40 weeks per year compared to a standard 52 weeks per year for full-time schedule.

Table 4. Estimated Parameters

Parameters	Values	Description	Target Moment	Data	Model
Fréchet					
η	0.5258	Hours Curvature	Earnings ratio Q4/Q2	1.270	1.271
$\underline{h}$	0.3999	Fixed Time Cost	Aggregate FT share (65+)	0.610	0.400
α	1.2828	Tail Behavior	Top curvature: Earnings ratio Q5/Q4	1.173	1.174
log-Gumbel					
η	1.256	Hours Curvature	Earnings ratio Q4/Q2	1.270	1.424
$\underline{h}$	0.030	Fixed Time Cost	Aggregate FT share (65+)	0.610	0.610
β	3.884	Tail Behavior	Top curvature: Earnings ratio Q5/Q4	1.173	1.171

Notes. All income distribution is divided into five quintile. Q1 denotes the lowest 20% and Q5 the highest 20% of income distribution. First row denotes the share of individuals (65 and older) who receive any earnings in 2017. Second row represents the share of earnings to the aggregate household income. Source: Figure 5 and Figure 6 in Congressional Research Service (2022)

However, the model captures the share of full-time workers across income quintile.

Given the long-standing debate on the structural parameter η, we study the model results with different values of η. This is shown in Figure A.3. Following the evidences from micro-side, we use $\eta = 0.2$ (MaCurdy, 1981; Altonji, 1986). This underestimates the share of full-time workers across income quintile. In contrast, if we use a parameter value $\eta = 1.0$ based on the evidences from macro-side (Chang and Kim 2006; Erosa et al. 2016), the model overestimates the share of full-time workers.

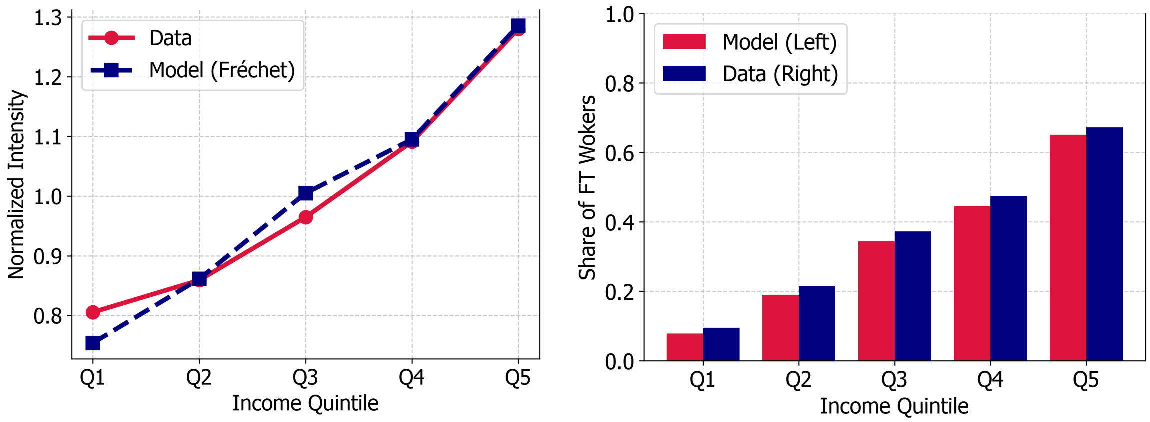
4.4 Decomposition of Aggregate Elasticity

We analyze a steday-state economy. Given the calibrated parameters, we simulate the labor supply response to a small perturbation in the marginal return to an hour. We then use the finite difference method to compute the Frisch elasticity for each income group. In this model, participation is endogenous, allowing for both an intensive and extensive margin response. We use the structural relation

$$\overline{z}_q = \Psi_q((1+\eta)\underline{h})^{1/\eta}$$

to define the participation threshold, which would be updated accordingly to changes in Ψ_q. The changes in aggregate hours is the sum of change in conditional hours worked

Figure 6. Model Fit

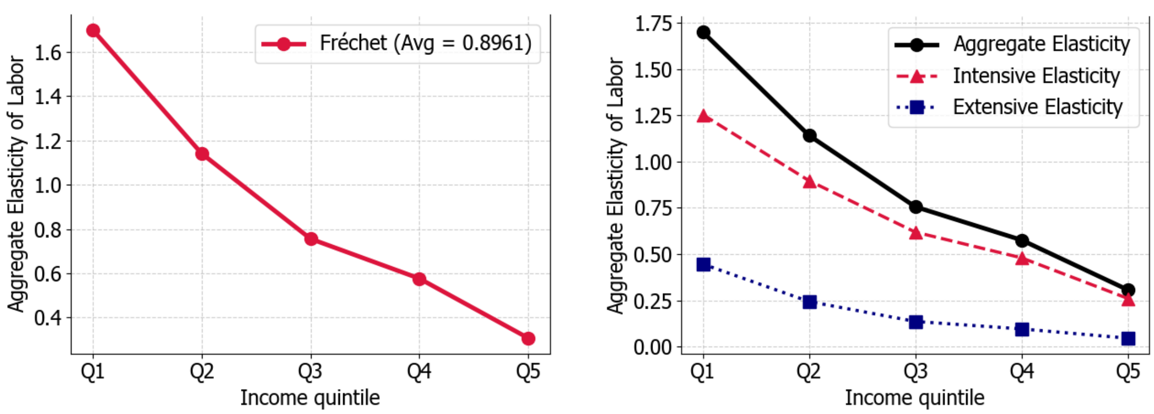


(a) Earnings Intensity (Intensive Margin)

(b) Share of FT Workers

Notes. For (a), the Fréchet distribution is used in estimating parameters. All the values are normalized by the average value. For (b), the share of full-time workers is compared with the data.

Figure 7. Aggregate Elasticity of Labor Supply



(a) Aggregate Elasticity of Labor Supply

(b) Decomposition of Aggregate Elasticity

Notes. For (a), the Fréchet distribution was used in estimating parameters. All the values are normalized by the average value. For (b), the share of full-time workers is compared with the values of p_q from the data.

weighted by employment rate, and the change in employment rate weighted by the average hours of those employed. Therefore, the aggregate Frisch elasticity for the economy is the employment-weighted average of these group-specific total elasticities.

Interpreting the results Figure 7 provides an aggregate Frisch elasticity and its decomposition. The left panel of Figure 7 shows the combined effect of both the intensive (hours) and extensive (participation) margins. The core finding is that the labor supply responsiveness is not uniform across the income distribution. Instead, it is monotonically decreasing with income.

The elasticity is at its peak for the lowest-income quintile (Q1), at a value of approximately 1.7. This indicates that a 1% change in the net wage (Ψ_1) would lead to 1.7% change in total hours worked for this group. This elasticity then falls sharply for each subsequent quintile, with 0.3 for the highest-income group (Q5). The average elasticity for the entire population is 0.896. This results is similar to 0.75 what Chetty et al. (2013) suggests for aggregate Frisch[5]

The right panel of Figure 7 decomposes the total response into two components: intensive and extensive elasticity. The result reveals two insights. First, for every income quintile, the intensive elasticity is significantly larger than the extensive elasticity. This means the primary way workers across the income distribution adjust to wage changes is by altering how much they work (e.g., moving from part-time to full-time or vice versa), rather than by moving in and out of the labor force. Second, both margins contribute to the overall downward slope, but the contribution share within each income quintile, the extensive margin has the largest contribution at the lowest-income group.

5 Counterfactuals

In this section, we anlayze two counterfactuals where the payroll tax rate ($\tau = 12.4\%$) is eliminated. First, we simulate the tax elimination using our baseline model. We measure the aggregate change in seniors' labor supply by income quintile, decomposing the effects along the intensive and extensive margins. Second, we then recalibrate the model to match the cross-sectional moments of prime-age workers. We run the same tax elimination experiment and compare their resulting labor supply change with that of the seniors.

5. Chetty et al. (2013) reports Frisch elasticities of 0.5 on the intensive margin and 0.25 on the extensive margin, for a combined elasticity of approximately 0.75 (Chetty et al. 2013, p. 26).

Table 5. Eliminating the Payroll Tax (τ = 12.4%): Working Hours and Full-Time Share Changes by Income

	Hours | Work			**FT Share (among employed)**		
Income bin	No Tax	Payroll Tax	Δ (%)	No Tax	Payroll Tax	Δ (%)
Q1	0.338	0.302	11.9	0.221	0.193	14.5
Q2	0.484	0.449	7.8	0.377	0.348	8.3
Q3	0.614	0.585	4.7	0.532	0.503	5.8
Q4	0.688	0.663	3.8	0.624	0.598	4.3
Q5	0.804	0.789	1.9	0.774	0.758	2.1
Aggregate	0.673	0.649	3.6	0.610	0.587	3.9

Notes: Hours|Work is measured as a fraction of working-hours out of full-time. FT Share is the fraction of employed who work full-time. Δ columns report percentage change (%) for Hours|Work and for FT Share. Aggregate is employment-weighted across quintiles. The Δ (%) columns calculate the percentage change from the "Payroll Tax" scenario to the "No Tax" scenario.

5.1 Eliminating Payroll Tax

Table 5 provides the intensive margin of labor supply - that is, the change in hours worked and full-time status among seniors who are already employed. The experiment simulates the complete elimination of the 12.4% payroll tax, which effectively acts as a wage increase for these workers. On aggregate, this policy induces a clear positive labor supply response. The average hours worked, measured as a fraction of a full-time schedule, increase by 3.6%. This increase is not just from a uniform rise in hours; it is also driven by a significant shift from part-time to full-time employment. The aggregate share of working seniors engaged in full-time work increases by 3.9%.

The most significant finding, however, is the heterogeneity of this response across the income distribution. The labor supply elasticity is highest for the lowest-income seniors and decliens monotonically as income rises. This is evident in both measures. For the intensive margins, the bottom quintile (Q1), whose baseline hours are the lowest (0.302), shows a large 11.9% increase in hours worked. In contrast, the top quintile (Q5), which already works the most (0.789), shows a minimal 1.9% increase.

This trend is even more pronounced in the full-time (FT) share. The baseline FT share for Q1 workers is very low at 19.3%, indicating they are predominantly part-time. When the tax is eliminated, this group's FT share increases by a substantial 14.5%. This suggests the tax is a major deterrent pushing these low-income workers to

Table 6. Eliminating the Payroll Tax ($\tau = 12.4\%$): Retirement and Employment Changes by Income

Income	Employment (No tax)	Employment (Payroll tax)	Δ Employment (pp)	Retiree as % of baseline workers
Q1	0.093	0.060	3.30	35.4
Q2	0.215	0.161	5.43	25.1
Q3	0.371	0.309	6.17	16.6
Q4	0.472	0.410	6.18	13.0
Q5	0.671	0.624	4.70	7.0
Aggregate	0.364	0.313	5.10	14.1

cap their hours. Conversely, the Q5 group, which is already 75.8% full-time, sees only a 2.1% increase. These results strongly imply that the labor supply of lower-income seniors, particularly their decision to move from part-time to full-time work, is far more sensitive to the net-of-tax wages than that of high-income seniors.

Table 6 provides the extensive margin of labor supply - the decision to work or retire. The results show that eliminating payroll tax has a substantial effect on the decision to work. In the baseline model with the tax, the employment rate for seniors is 31.3%. When the tax is removed, the employment rate jumps to 36.4%. This is an increase of 5.10 percentage points (pp), which a big response. The final column of Table 6 make up 14.1% of the total workforce in the baseline model, indicating a large influx of previously-retired seniors into the labor market.

The changes of employment in percentage points shows a inverted U-shape pattern. The lowest-income group has the smallest response followed by the highest-income group. The strongest increases in the employment comes from the middle-income groups. However, the final column of Table 6 shows how much the composition of the workforce changes. For the Q1 group, while the total number is small (3.3pp), they are a large part of the new Q1 workforce. Over one-third or 35.4% of all working seniors in Q1 in the baseline model are new entrants. This means the policy dramatically changes the composition of the low-income senior workforce. Conversely, for the Q5 group, the workforce is highly stable. The new workers incentivized by the policy change (4.7pp) make up only 7% of the total Q5 workforce. This is because most of the Q5 workers were already employed in the baseline.

The elimination of payroll tax has different implications for income quintile. For the lowest-income seniors, the payroll tax is not the main driver of their retirement decision. This may due to poor health or low wage potentials. But it is a big determinant of how much they work, acting as a major incentive to remain in part-time employment. The middle-income seniors are the most on the margin. The payroll tax rate is the critical factor in the retirement decision itself. Removing it incentivizes the largest number of them back in the workforce. The high-income seniors are less responsive on the extensive margin compared to the middle-income group.

5.2 Seniors vs. Prime-Age Workers

Table 7 shows the effect of payroll tax elimination on the intensive margins of labor supply for prime-age workers. On aggregate, prime-age workers respond to the tax cut by working more. Average hours worked increase by 2.1%, and the share of who are full-time increases by 2.49%. This is a standard substitution effect. Similarly, important findings of the result is the clear heterogeneity across income groups. The labor supply response is strongest for the lowest-income quintile and declines monotonically as income rises.

In aggregate, eliminating the payroll tax increases the labor supply of prime-age workers by 2.1% while seniors' labor supply increase by 3.6%. The ratio of these responses is 1.71, suggesting that seniors in our model are 1.71 times as elastic as prime-age workers.

We acknowledge that because our model is a partial equilibrium, it does not capture general equilibrium effects. Therefore, these aggregate figures should not be taken at face value as a measure of the total macroeconomic response. However, the main objective of this exercise is a comparative one. Given that the only difference between the two counterfactuals is the set of cross-sectional moments used for calibration, this relative comparison of labor supply responses between seniors and prime-age workers remain plausible and informative.

Table 7. Eliminating the Payroll Tax ($\tau = 12.4\%$): Prime-age Workers

	Hours | Work			**FT Share (among employed)**		
Income Bin	No Tax	Payroll Tax	Δ (%)	No Tax	Payroll Tax	Δ (%)
Q1	0.8650	0.8403	2.94	0.8491	0.8203	3.51
Q2	0.8687	0.8447	2.84	0.8561	0.8278	3.42
Q3	0.8776	0.8573	2.37	0.8744	0.8514	2.70
Q4	0.8927	0.8771	1.78	0.9026	0.8835	2.16
Q5	0.9080	0.8967	1.26	0.9317	0.9174	1.56
Aggregate	0.8856	0.8674	2.10	0.8889	0.8673	2.49

Notes. Hours|Work is measured as a fraction of working-hours out of full-time. FT Share is the fraction of employed who work full-time. Δ columns report percentage change (%) for Hours|Work and for FT Share. Aggregate is employment-weighted across quintiles. The Δ (%) columns calculate the percentage change from the "Payroll Tax" scenario to the "No Tax" scenario.

6 Conclusion

Ensuring the long-term solvency of the Social Security system is a critical policy challenge, particularly as projections indicate its trust funds will be depleted within the next decade, necessitating significant reform. Since retirement choices are intricately linked to Social Security, a complete analysis requires understanding how the payroll tax affects the retirement of the seniors. Ultimately, the choice to retire is a fundamental labor supply decision, wherein senior workers compare the marginal benefits of continued employment against the associated costs.

We develop a tractable labor supply model to explain a key feature of the U.S. labor market of the seniors: high-income seniors tend to work more, and for longer hours. The novelty of the model is to account for the large variation in an individual costs and disincentives of working based on Extreme Value distributions. After calibrating the model to match U.S. economic data in 2017 and 2018, the model replicates an empirical observation that as income rises, so do hours worked and the share of working full-time of the seniors.

Using our model as a simulated laboratory, we examine how eliminating the payroll tax affects seniors' labor supply. Our central findings are twofold. First, the effects are heterogeneous across the income distribution. We find that low- and middle-income seniors are significantly affected by the reform, while the impact on the labor decisions of high-income seniors is relatively small. Second, our results show that seniors are 1.71 times more elastic than prime-age workers.

These findings answer a simple but policy-relevant quesiton: Would increasing the payroll tax rate resolve the Social Security's solvency? Our results suggest the opposite. An increase in the payroll tax would induce the most elastic groups - low and middle income seniors - to exit the labor force and claim Social Security benefits early. With fewer workers paying in and more retirees drawing out, this would perversely worsen, not improve, the financial pressure on Social Security.

References

Abramson, B., Boerma, J., and Tsyvinski, A. (2024). Macroeconomics of Mental Health. Technical Report w32354, National Bureau of Economic Research.

Aguiar, M., Bils, M., Charles, K. K., and Hurst, E. (2021). Leisure Luxuries and the Labor Supply of Young Men. *Journal of Political Economy*, 129(2):337–382.

Aguiar, M. and Hurst, E. (2016). Chapter 4 - The Macroeconomics of Time Allocation. In Taylor, J. B. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2, pages 203–253. Elsevier.

Ameriks, J., Briggs, J., Caplin, A., Lee, M., Shapiro, M. D., and Tonetti, C. (2020). Older Americans Would Work Longer If Jobs Were Flexible. *American Economic Journal: Macroeconomics*, 12(1):174–209.

Bick, A., Blandin, A., and Rogerson, R. (2022). Hours and Wages. *The Quarterly Journal of Economics*, 137(3):1901–1962.

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

Chetty, R., Guren, A., Manoli, D., and Weber, A. (2011). Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins. *American Economic Review*, 101(3):471–475.

Chetty, R., Guren, A., Manoli, D., and Weber, A. (2013). Does Indivisible Labor Explain the Difference between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities. *NBER Macroeconomics Annual*, 27:1–56.

Coile, C. C. (2025). 12 Changing Retirement Incentives and Retirement in the US. In *Social Security Programs and Retirement around the World*, chapter Social Security Programs and Retirement around the World, pages 401–418. University of Chicago Press.

Congressional Research Service (2022). Income for the Population Ages 65 and Older: Evidence from the Health Retirement Study (HRS). Legislation R47341, Congressional Research Service.

Duggan, M., Dushi, I., Jeong, S., and Li, G. (2023). The effects of changes in social security's delayed retirement credit: Evidence from administrative data. *Journal of Public Economics*, 223:104899.

Erosa, A., Fuster, L., and Kambourov, G. (2016). Towards a Micro-Founded Theory of Aggregate Labour Supply. *The Review of Economic Studies*, 83(3):1001–1039.

Ferraro, D. and Valaitis, V. (2024). Consumption Quality and Employment Across the Wealth Distribution. *The Review of Economic Studies*, page rdae052.

French, E. (2005). The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour. *The Review of Economic Studies*, 72(2):395–427.

Gustman, A. L. and Steinmeier, T. L. (1986). A Structural Retirement Model. *Econometrica*, 54(3):555–584.

Hansen, G. D. (1985). Indivisible labor and the business cycle. *Journal of Monetary Economics*, 16(3):309–327.

Imrohoroglu, S. and Kitao, S. (2012). Social Security Reforms: Benefit Claiming, Labor Force Participation, and Long-run Sustainability. *American Economic Journal: Macroeconomics*, 4(3):96–127.

Kaplan, G. and Schulhofer-Wohl, S. (2018). The Changing (Dis-)utility of Work. *Journal of Economic Perspectives*, 32(3):239–258.

Karabarbounis, M. (2016). A Road Map for Efficiently Taxing Heterogeneous Agents. *American Economic Journal: Macroeconomics*, 8(2):182–214.

Kleven, H., Kreiner, C., Larsen, K., and Søgaard, J. (2025). Micro versus Macro Labor Supply Elasticities: The Role of Dynamic Returns to Effort. *American Economic Review*, 115(9):2849–2890.

Ljungqvist, L. and Sargent, T. J. (2011). A Labor Supply Elasticity Accord? *The American Economic Review*, 101(3):487–491.

Ljungqvist, L. and Sargent, T. J. (2014). Career length: Effects of curvature of earnings profiles, earnings shocks, taxes, and social security. *Review of Economic Dynamics*, 17(1):1–20.

Maestas, N. (2010). Back to Work: Expectations and Realizations of Work after Retirement. *The Journal of Human Resources*, 45(3):718–748.

Nakajima, M. and Telyukova, I. A. (2025). Medical Expenses and Saving in Retirement: The Case of the United States and Sweden. *American Economic Journal: Macroeconomics*, 17(1):161–202.

Prescott, E. C. (2004). Why Do Americans Work So Much More Than Europeans? *Quarterly Review*, 28(July).

Rogerson, R. (1988). Indivisible labor, lotteries and equilibrium. *Journal of Monetary Economics*, 21(1):3–16.

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

Rogerson, R. and Wallenius, J. (2013). Nonconvexities, Retirement, and the Elasticity of Labor Supply. *American Economic Review*, 103(4):1445–1462.

Shoven, J., Slavov, S. N., and Wise, D. (2017). Social Security Claiming Decisions: Survey Evidence. Technical Report w23729, National Bureau of Economic Research, Cambridge, MA.

Stock, J. H. and Wise, D. A. (1990). Pensions, the Option Value of Work, and Retirement.

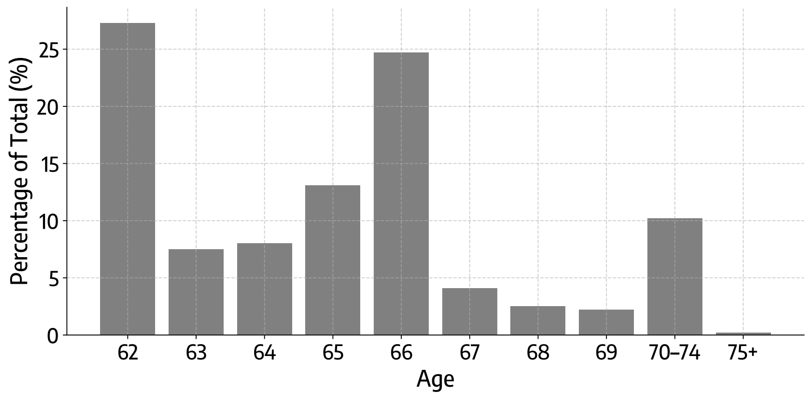
Econometrica, 58(5):1151–1180.

Yu, Z. (2024). Why are older men working more? The role of social security. *Journal of Public Economics*, 231:105071.

Appendices

A Figures

FIGURE A.1. Age Distribution of Claiming Social Security Benefits



Notes. American retired workers who newly claimed Social Security in 2022. Data source: Annual Statistical Supplement to the Social Security Bulletin, 2023

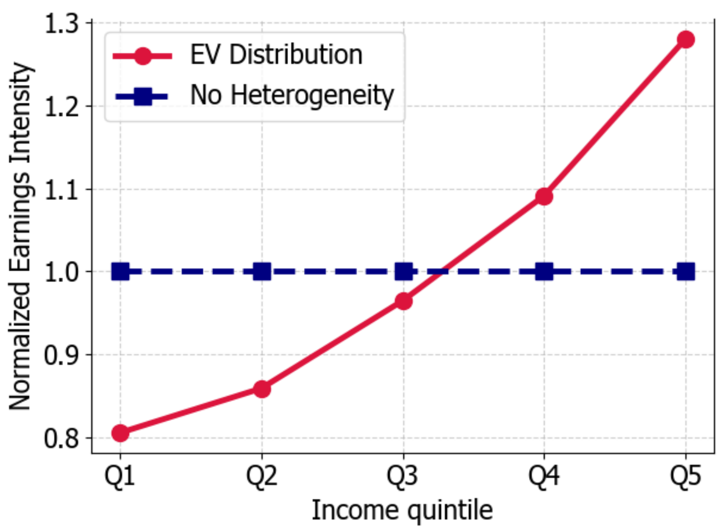
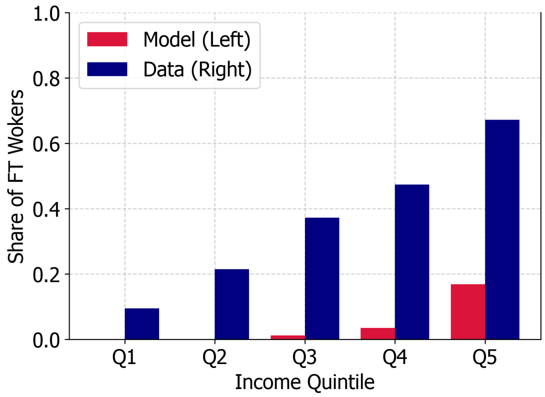


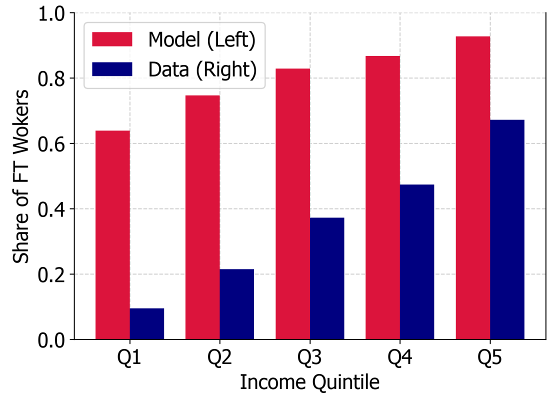
FIGURE A.2. Role of Heterogeneity

Notes. The Fréchet distribution was used in estimating parameters with shape parameter α. For no heterogeneity case, we assume a limit case of Fréchet distribution, $\alpha \to \infty$, which makes the distribution degenerate. All the values are normalized by the average value.

Figure A.3. Different Structural Parameters η



(a) $\eta = 0.2$



(b) $\eta \geq 1.0$

Notes. For (a), (MaCurdy, 1981; Altonji, 1986). For (b), (Chang and Kim 2006; Erosa et al. 2016)

Table A.1. Target Moments by Income Quintile (Prime-age Workers)

Income Quintile	**Q1**	**Q2**	**Q3**	**Q4**	**Q5**
Share of individuals with earnings	41.1%	42.9%	50.2%	58.8%	69.1%
Share of earnings to total income	44%	66%	79%	87%	96%
Share of Full-Time Employment	88.9%				

Notes. All income distribution is divided into five quintile. Q1 denotes the lowest 20% and Q5 the highest 20% of income distribution. First row denotes the share of individuals (65 and older) who receive any earnings in 2017. Second row represents the share of earnings to the aggregate household income. Source: Figure 5 and Figure 6 in Congressional Research Service (2022)

Table A.2. Estimated Parameters (Prime-age Workers)

Parameters	**Values**	**Description**	**Target Moment**	**Data**	**Model**
Fréchet					
η	1.999	Hours Curvature	Earnings ratio Q4/Q2	0.961	1.012
$\underline{h}$	0.278	Fixed Time Cost	Aggregate FT share (65+)	0.889	0.889
α	3.999	Tail Behavior	Top curvature: Earnings ratio Q5/Q4	0.939	1.007
log-Gumbel					
η	1.999	Hours Curvature	Earnings ratio Q4/Q2	0.961	1.028
$\underline{h}$	0.073	Fixed Time Cost	Aggregate FT share (65+)	0.889	0.889
β	3.999	Tail Behavior	Top curvature: Earnings ratio Q5/Q4	0.939	1.017

Table A.3. Model Fit: Normalized Intensity Shape by Quintile

Quintile	Data (normalized intensity)	Model (Fréchet)	Model (Gumbel log)
Q1	0.759	0.976	0.975
Q2	1.090	0.980	0.980
Q3	1.115	0.994	0.994
Q4	1.049	1.013	1.014
Q5	0.985	1.034	1.035

B Appendix - Derivation

B.1 Deriving Aggregate Frisch Elasticity

Here we lay out how to derive equation (9). Note that the aggregate labor supply is written as

$$H_t = F(\underline{z}(t)) + \Psi_t^\eta \int_{\underline{z}(t)}^{\overline{z}(t)} z^{-\eta} dF(z) - \underline{h}F(\overline{z}(t))$$

Differentiate each terms with respect to marginal return to an hour $\ln \Psi_t$. For the integral term, we apply Leibniz's rule to

$$I(\Psi_t) = \int_{\underline{z}(t)}^{\overline{z}(t)} z^{-\eta} dF(z).$$

Because the term $z^{-\eta}$ is independent of Ψ_t, we have

$$\begin{aligned}
\frac{dI}{d\ln\Psi_t} &= \overline{z}^{-\eta} f_Z(\overline{z}) \frac{d\overline{z}}{d\ln\Psi_t} - \underline{z}^{-\eta} f_Z(\underline{z}) \frac{d\underline{z}}{d\ln\Psi_t} \\
&= \overline{z}^{1-\eta} f_Z(\overline{z}) - \underline{z}^{1-\eta} f_Z(\underline{z}).
\end{aligned}$$

Now differentiate the product $\Psi_t I(\Psi_t)$ gives us

$$\frac{d}{d\ln(\Psi_t)}[\Psi_t I(\Psi_t)] = \eta\Psi_t^\eta I(\Psi_t) + \Psi^\eta \Big(\overline{z}^{1-\eta} f_Z(\overline{z}) - \underline{z}^{1-\eta} f_Z(\underline{z})\Big).$$

Putting all these terms together, we have

$$\begin{aligned}
\frac{dH}{d\ln(\Psi_t)} &= \underline{z} f_Z(\underline{z}) - \underline{h}\overline{z} f_Z(\overline{z}) + \eta\Psi_t^\eta \int_{\underline{z}(t)}^{\overline{z}(t)} z^{-\eta} dF(z) + \Psi^\eta \overline{z}^{1-\eta} f_Z(\overline{z}) - \underline{z}^\eta \underline{z}^{1-\eta} f_Z(\underline{z}) \\
&= \Psi^\eta \overline{z}^{1-\eta} f_Z(\overline{z}) + \eta\Psi_t^\eta \int_{\underline{z}(t)}^{\overline{z}(t)} z^{-\eta} dF(z) - \underline{h}\overline{z} f_Z(\overline{z}).
\end{aligned}$$

From the threshold rule at the extensive margin,

$$\frac{1}{1+\eta} \frac{\Psi_t^{1+\eta}}{\overline{z}^\eta} = \Psi_t \underline{h} \Longrightarrow \Psi_t^\eta \overline{z}^{-\eta} = (1+\eta)\underline{h}$$

Multiplying $\overline{z}$ on both sides,

$$\Psi_t^\eta \overline{z}^{1-\eta} = (1+\eta)\overline{z}\underline{h}$$

and substituting this to the above equation, we get

$$-\underline{h}\overline{z}f_Z(\overline{z}) + \Psi^{\eta}\overline{z}^{1-\eta}f_Z(\overline{z}) = -\underline{h}\overline{z}f_Z(\overline{z}) + (1+\eta)\overline{z}\underline{h} = \eta\underline{h}\overline{z}f_Z(\overline{z})$$

Then we get the aggregate Frisch elasticity (9),

$$\frac{d\ln(H)}{d\ln(\Psi_t)} = \frac{\eta}{H}\Psi_t^{\eta}\int_{\underline{z}(t)}^{\overline{z}(t)} z^{-\eta}dF(z) + \eta\underline{h}\overline{z}f_Z(\overline{z}).$$

To complete the derivation, we turn our focus to deriving individual Frisch elasticity. Consider only the part-time region of working $z \in [\underline{z}, \overline{z}]$. Individual Frisch elasticity is

$$\varepsilon_{\text{indiv}}(z) \equiv \frac{d\ln(h)}{d\ln(\Psi_t)} = \eta\frac{h(z)+\underline{h}}{h(z)}.$$

Multiply $h_t(z)$ on both sides,

$$\varepsilon_{\text{indiv}}(z)h(z) = \eta(h(z)+\underline{h}).$$

Using the optimality condition of labor supply,

$$h + \underline{h} = \left(\frac{\Psi_t}{z}\right)^{\eta}$$

which derives the full aggregate Frisch elasticity

$$\frac{d\ln(H_t)}{d\ln(A_t)} = \underbrace{\frac{1}{H}\int_{\underline{z}(t)}^{\overline{z}(t)} \varepsilon_{\text{indiv}}(z)h(z)dF_Z(z)}_{\text{Intensive margin}} + \underbrace{\frac{\eta\underline{h}}{H_t}\,\overline{z}(t)f_Z(\overline{z}(t))}_{\text{Extensive margin}}.$$

B.2 Indivisible Labor Supply Model

We use the time-averaging model of Ljungqvist and Sargent (2014) with indivisible labor. The planner achieves the same feasible set and optimality conditions as in Rogerson (1988) when the planner's subjective discount rate equals interest rate in a life-cycle setting. In other words, we can interpret H_t as either a cross-sectional employment rate, or a time share of employment throughout life-cycle. Indivisible choices at the individual level aggregate into a divisible object H_t.

Consider a life-cycle model where time is normalized to $t \in [0,1]$ with indivisible labor $h_{jt} \in \{0,1\}$. Assume there is a unit continuum of households indexed by $j \in [0,1]$, who draw an effective disutility of working Z with a cdf F_Z and density function f_z on its support. Since labor supply is indivisible, B is an average disutility of work. An idiosyncratic taste shocks are independent and identically distributed (iid) across (j, t) with a cumulative distribution function F with density f positive on support. Assume a technology that is linear in labor, or wH, with wage $w > 0$ and interest rate r are given. The planner discounts at ρ.

The planner chooses $\{c_{jt}, h_{jt}\}$ to

$$\max_{\{c_{jt}(z), h_{jt}(z)\}} \int_0^1 e^{-\rho t} \left[\int_0^1 \left(u(c_{jt}(z)) - \int z h_{jt}(z) dF_Z(z) \right) dj \right] dt \tag{.1}$$

subject to a resource constraint

$$\int_0^1 e^{-rt} \left[\int_0^1 \left(\int ((1-\tau) w h_{jt}(z) - c_{jt}(Z) + T_t) dF_Z(z) \right) dj \right] dt \geq 0 \tag{.2}$$

where all the taxes paid are rebated as a lump-sum transfer T_t. With complete markets, the optimal consumption is equalized across households within each period t and shock z , so that $c_{jt}(z) = c^*$ for all (j, t, z). Since the planner sorts households by the value of Z, there exists a threshold $z^*(t)$ with

$$h_t(z) = \mathbf{1}\left(z \leq z^*(t)\right)$$

which is an indicator function for the argument. Define the total employment rate

$$H_t = \int h_t(z) dF_Z(z) = F_Z(z^*).$$

Then the total disutility of employment at time t is

$$\int z h_t(z) dF(z) = \int_0^{H_t} F_Z^{-1}(q) dq = \Psi(H_t) \tag{.3}$$

where I use change-of-variables to map the threshold rule to total disutility Ψ. The boundary conditions are $\Psi(0) = 0$ and $\Psi(1) = B = \mathbb{E}[Z]$. Note that $\Psi'(H) = F_Z^{-1}(H) \geq 0$ for all H on $Z \in [0, \infty)$, and $\Psi''(H) = \frac{1}{f_Z(F_Z^{-1}(H))} > 0$ for each support, implying the cost function $\Psi(H)$ is convex. Since the utility function is concave by assumption, the planner has a concave objective function with full support on Z, which leads to a unique solution of H^*.

Consider a unit mass of ex-ante identical hosueholds with indivisible hours $h \in$

$\{0, 1\}$ and per-period utility $u(c) - v(h)$ with an average disutility of work $v(1) = B$. Idiosyncratic taste shocks or work-costs are drawn from Type-1 Extreme Value distribution with a shape parameter μ. Let $H_t \in [0, 1]$ denote the fraction of employed at time t, and c_t denote consumption. When the subjective discount rate equals to interest rate ($\rho = r$), the planner's problem is

$$\max_{c_t, H_t} \int_0^1 e^{-\rho t} \Big(u(c_t) - \Psi(H_t) \Big) \, dt \tag{.4}$$

subject to $c_t \leq (1 - \tau) w_t H_t + T_t$. A convex employment cost can be written in a closed-form for an i.i.d taste shocks as in (.3),

$$\begin{aligned} \Psi(H) &= \int_0^H \left[B + \mu \log\left(\frac{q}{1-q} \right) \right] dq \\ &= BH + \mu \left[H \log H + (1 - H) \log(1 - H) \right] \end{aligned}$$

where the first line of equation uses a functional form $F_{\Delta\xi}(x) = 1/(1 + e^{-x/\mu})$ for $H \in [0, 1]$. For the second line, we use the definition of inverse cdf property. Note that if there is no idiosyncratic heterogeneity or $\mu \to 0$, the total cost of employment is BH which is the same as in Rogerson (1988) and there exists a representative agent with linear labor, an important aspect of indivisible labor in producing high aggregate elasticity of labor supply.