When in Doubt, Tax More Progressively? Uncertainty and Progressive Income Taxation*

Minsu Chang† Chunzan Wu‡

May 10, 2023

Abstract

We study the optimal income tax problem when policymakers have only limited information about household preferences and wage process. We provide conditions that qualitatively determine the impact of such parameter uncertainty on optimal tax policy. To quantify this effect, we build and estimate an incomplete-market life-cycle model of heterogeneous households using a limited-information Bayesian approach with U.S. data. We find that such uncertainty leads to a more progressive optimal income tax policy, resulting in a 5 percentage point increase in the marginal tax rate gap between high- and low-income households. This result is primarily driven by uncertainty about wage process, and the correlations between uncertain parameters and the shape of their posterior distribution play an important role.

JEL Codes: E60, H20.
Keywords: Parameter Uncertainty, Progressive Tax, Optimal Taxation.

*We thank Dan Cao, Behzad Diba, Jesús Fernández-Villaverde, Jonathan Heathcote, Mark Huggett, Dirk Krueger, Toshihiko Mukoyama, Gaston Navarro, and Margit Reischer for helpful comments and suggestions.

†Georgetown University. Email: minsu.chang@georgetown.edu.
‡Peking University. Email: czwu@nsd.pku.edu.cn.

1 Introduction

How should the government tax people's income? This is one of, if not the, most important questions faced by policymakers. Tremendous efforts of the economics profession, both in micro and macroeconomics, through theoretical and quantitative approaches, have been devoted to answering this question; and thanks to that, substantial progress has been made on this front. For example, the previous literature has identified several primitives crucial to the optimal design of income tax such as the elasticity of labor supply and the amount of uninsurable idiosyncratic risk in the economy. However, policymakers only have limited information about these primitives because they are not observable and must be inferred from the data. How does such imperfect information alter our answer to the optimal tax question?

In this paper, we study the income tax policy implications of information frictions in the form of policymakers' uncertainty about parameters governing household preferences and wage process. To fix ideas, we first derive general conditions that determine qualitatively how parameter uncertainty affects the optimal tax policy.[1] We show that the effect of parameter uncertainty depends on the convexity of the welfare measure with respect to the uncertain parameter, and how this convexity interacts with tax policy. Intuitively, policymakers can improve welfare by leveraging the relationship between tax policy and the convexity of the welfare measure. In addition, when there are multiple uncertain parameters, we show that the correlations between them also matter.

To quantify the effects of uncertainty, we build an incomplete-market life-cycle model of heterogeneous households with idiosyncratic wage risk and endogenous labor supply. This model serves two goals. First, through the lens of this model, we interpret the data and assess the uncertainty about key economic parameters. Second, it provides us with a laboratory for conducting thought experiments of tax reforms and evaluating their welfare consequences. We extend the standard

[1] It is worth clarifying that parameter uncertainty here means the dispersion of policymakers' belief about parameters, holding constant their expected values. And even if the ex post optimal tax policy varies strongly with a parameter, it does not necessarily imply that the uncertainty about this parameter is also important to the ex ante optimal tax policy because the effects of possible higher and lower parameter values might offset each other.

single-earner household model by incorporating explicitly the endogenous labor supply decisions of the secondary earners (typically females), since they serve as an important private insurance mechanism and have major implications for income tax policy.

The severity of information frictions is measured by the degree of policymakers' uncertainty about parameters governing household preferences and wage process (e.g., household risk aversion and labor supply elasticities; persistences and variances of wage shocks), which in turn depends on how much information about these parameters can be extracted from the data. To assess the amount of uncertainty, we estimate these parameters through a limited-information Bayesian approach, using the data from the 1999-2017 Panel Study of Income Dynamics (PSID).[2] The Bayesian inference produces a joint posterior distribution of preference and wage parameters that summarizes policymakers' uncertainty about these aspects of the economy.

To study the tax policy implications of uncertainty, as represented by the posterior distribution of parameters, we conduct thought experiments with our life-cycle model, in which policymakers choose the income tax to maximize the welfare of a newborn cohort subject to a within-cohort budget constraint.[3] Following Bénabou (2002) and Heathcote et al. (2017), the income tax policy is summarized by two parameters of a nonlinear tax function $\widetilde{T}(Y) = Y - (1-\chi)Y^{1-\mu}$, where Y is household income, $\widetilde{T}(Y)$ is tax liability, and μ and χ represent the progressivity and the level of income tax, respectively.

Policymakers face the classic equity-efficiency trade-off when choosing the tax policy. On the one hand, a progressive income tax reduces inequality and provides valuable social insurance against idiosyncratic wage risks, which are not fully insurable among households because the financial markets are incomplete. On the other hand, a progressive income tax implies a rising marginal tax rate with in-

[2]The limited-information Bayesian approach can be viewed as the Bayesian version of the generalized method of moments (GMM). It enables us to i) quantify parameter uncertainty as the posterior distribution that the policymakers' objective function is based on, and ii) avoid the forbidding task of constructing the full likelihood function for a complex nonlinear model like ours.

[3]We confine the optimal tax policy problem to a newborn cohort to avoid the redistribution motive of income tax over different generations (and the tricky question of how to weight them).

come, which discourages household labor supply and leads to efficiency losses. Both the benefits and costs of progressive income tax are regulated by the underlying preference and wage parameters, and so is the optimal tax policy. However, when policymakers are uncertain about the values of these parameters, there is a new dimension of policy concern: that is, the policymakers must also balance the performance of tax policy and trade off welfare gains and losses across scenarios with different parameter values.

The main finding of our quantitative analysis is that taking into account the uncertainty about preference and wage parameters leads to more progressive optimal income tax. We reach this conclusion by contrasting the results from two scenarios. In the first scenario, as conventionally assumed in the literature, policymakers ignore parameter uncertainty and search for the optimal policy treating the point estimates (i.e., expected values) of parameters as their true values. In the alternative scenario, policymakers reflect parameter uncertainty by maximizing the expected welfare with respect to the posterior distribution.[4] We find that the optimal tax progressivity is about 2 percentage point higher once parameter uncertainty is accounted for ($\mu = 0.13$ vs. 0.15), which translates to a roughly 5 percentage point increase in the marginal tax rate faced by the high-income households relative to those near the bottom of the income distribution.

Through decomposition exercises, we find that this increase in optimal tax progressivity is primarily driven by uncertainty about wage parameters (112%), whereas the effect of uncertainty about preference parameters is much smaller and in the opposite direction (-12%). The larger effect of uncertainty about wage parameters is not only because there is greater uncertainty about these parameters, but also because the (ex ante) optimal tax progressivity is more sensitive to such uncertainty. Similarly, we find that a majority of the effect on optimal tax progressivity is from uncertainty about male-related parameters (90%), followed by female-related (19%) and gender-neutral (-9%) parameters. Although uncertainty about female-related parameters has an overall minor effect, the optimal tax pro-

[4] We assume that policymakers take parameter uncertainty as exogenous when choosing the tax policy. Allowing policy decisions to potentially influence the information received by policymakers (and thus the uncertainty and policy in the future) is an interesting direction for future research.

gressivity is actually most sensitive to such uncertainty.

We estimate the posterior distribution from the data, which contains information about the correlations between uncertain parameters and the shape of their distribution. Through counterfactual experiments, we also demonstrate that this information is crucial for quantitative evaluations about the effects of parameter uncertainty. For example, reversing the correlations between uncertain parameters or assuming a uniform posterior distribution could lead to substantially different results about the effect of uncertainty: the change in optimal tax progressivity is 25% smaller or 50% larger compared to that based on the true posterior distribution.

We also show, through quantitative analysis, that the welfare cost of uncertainty about preference and wage parameters is about 0.2% of household lifetime consumption (approximately 29 billion dollars per year based on 2019 U.S. consumption data) *through the income tax policy channel alone*. To measure the welfare cost of parameter uncertainty, we compute the amount of welfare the policymakers are just willing to give up in exchange for a signal that reveals the true values of parameters. Because in addition to the income tax policy, there are also other government policy decisions that rely on the information about household preferences and wage process, the total welfare cost of this uncertainty through distortions to all policy decisions is likely much larger. Our finding thus indicates substantial potential welfare gains from activities that improve our knowledge about these aspects of the economy.

Our benchmark analysis assumes that policymakers are risk-neutral with regard to variations in the welfare consequence of tax reform induced by parameter uncertainty. For completeness, we also investigate how alternative risk preferences of policymakers affect our finding about optimal tax progressivity. We consider two types of policymakers: i) "career politicians" who maximize the welfare gain relative to the status quo, and ii) "social planners" who minimize the welfare loss relative to the first-best scenario with perfect information about parameters.[5] Interestingly, we find that for career politicians, the optimal tax progressivity is reduced by their risk aversion, whereas for social planners, the optimal tax progressivity re-

[5]The two types of policymakers are equivalent when they are risk-neutral, and hence we do not need to differentiate them for the benchmark results.

mains almost the same. Intuitively, for career politicians, maintaining the status quo policy is risk-free since the welfare change is zero independent of parameter values. Therefore, the risk aversion of career politicians pushes the optimal income tax closer to the status quo, which means lower tax progressivity compared to the risk-neutral case. For social planners, a risk-free policy must be parameter-contingent, which is not feasible facing parameter uncertainty. For example, the status quo policy is still risky for social planners since the welfare loss of inaction relative to the first-best is uncertain. Put another way, by doing nothing, social planners could be giving up very little or a lot in potential welfare gain, depending on the true state of parameters. We find that the optimal tax progressivity is fairly robust with respect to social planners' risk aversion.

This paper is most related to the literature that studies the optimal nonlinear income tax in the Ramsey tradition with heterogeneous-agent incomplete-market models. Influential previous studies include Bénabou (2002), Conesa and Krueger (2006), Conesa et al. (2009), and more recently Bakış et al. (2015), Krueger and Ludwig (2016), and Heathcote et al. (2017) among others. A common implicit assumption of these previous studies is that, when choosing the tax policy, policymakers have perfect information about the parameters of their model for the economy. This paper relaxes that assumption and incorporates policymakers' concern about parameter uncertainty into the policymaking process.

The way that parameter uncertainty is analyzed in this paper has several advantages compared to the sensitivity analysis in the literature, in which quantitative exercises are simply repeated for a list of alternative calibrations of parameters. One that is particularly important for optimal policy studies is that sensitivity analysis only produces a parameter-contingent policy plan, which is not implementable when policymakers don't know the true values of parameters.[6] Our approach also differs from the robust control literature as represented by Hansen and Sargent

[6] Additionally, parameter values in sensitivity analysis are often drawn from previous empirical studies, which might be estimated under alternative assumptions that are inconsistent with the model for policy analysis. In contrast, we infer parameter uncertainty from the data through the lens of the same model for policy analysis. The posterior distribution of parameters also carries information about the correlations between parameters, which we show are important for both the qualitative and quantitative effects of uncertainty (Section 2.1 and 5.1.4).

(2008). Their approach aims at finding a policy that performs well across alternative models without information on the plausibility of each alternative model, and hence the optimal policy becomes the solution to a max-min problem to avoid worst-case scenarios. In contrast, we analyze model uncertainty (as represented by the values of parameters) from a Bayesian perspective and infer the probabilities of alternative models from the data, which are then used to weight different scenarios in the policymaking process à la Brainard (1967).[7]

Earlier research on policy implications of parameter uncertainty focuses primarily on monetary policy. Representative quantitative works include Levin et al. (2005) and Edge et al. (2010), which study the optimal monetary policy under parameter uncertainty in micro-founded New Keynesian models. In comparison, studies of the tax policy implications of parameter uncertainty are rare and mostly theoretical. Lockwood et al. (2020) studies the optimal income tax problem with uncertainty over the elasticity of taxable income in the Mirrleesian tradition. They consider a simple static environment for tractability, whereas our quantitative dynamic model includes additional features of the economy that are important for tax policy evaluations such as household life cycle and endogenous private insurance through precautionary savings and family labor supply. We also quantify the welfare cost of parameter uncertainty and consider the effects of risk preferences of policymakers.

The rest of this paper proceeds as follows. Section 2 derives qualitative theoretical results about parameter uncertainty's effect on optimal tax policy. Section 3 sets up the incomplete-market life-cycle model for our quantitative analysis. Section 4 describes our empirical strategy for estimating model parameters and reports the parameter uncertainty inferred from the data. Section 5 conducts thought experiments of tax reforms and quantifies the tax policy implications of parameter uncertainty. Section 6 concludes.

2 Theoretical Analysis

In this section, we derive general conditions that determine qualitatively how parameter uncertainty affects the optimal design of tax policy. We then illustrate

[7] We do consider a special case in Section 5.3 that is similar in spirit to the max-min criterion.

the use of these conditions in a tractable static model, focusing on uncertainty about the elasticity of labor supply and the magnitude of idiosyncratic wage risk.

2.1 General Analysis

Without loss of generality, suppose that the tax policy is summarized by a single policy parameter μ that represents the progressivity of the tax schedule, and there is a single economic parameter θ that the policymakers are uncertain about. Let $W(\mu, \theta)$ denote the welfare gain from a tax reform that switches from the status quo policy to tax policy μ if the true value of the economic parameter is θ.

Policymakers have partial information about the economic parameter θ, and their posterior belief about its value is summarized by a probability distribution with the cdf $F_\theta(\theta)$. When uncertainty is ignored, policymakers treat the point estimate (i.e., expected value) of θ, $\bar{\theta} = \int \theta dF_\theta(\theta)$, as its true value and choose the tax policy $\bar{\mu}$ to maximize the welfare gain $W(\mu, \bar{\theta})$:

$$\bar{\mu} = \arg\max_{\mu} W(\mu, \bar{\theta}).$$

When uncertainty is accounted for, policymakers should choose the optimal policy μ^* that maximizes the expected welfare gain $\widetilde{W}(\mu)$ based on the posterior distribution:

$$\mu^* = \arg\max_{\mu} \underbrace{\int_\theta W(\mu, \theta) dF_\theta(\theta)}_{\equiv \widetilde{W}(\mu)},$$

where policymakers are assumed to be risk-neutral with regard to the welfare risk induced by parameter uncertainty.

The question of interest here is how taking into account the uncertainty about θ affects the optimal tax policy, i.e., how different μ^* is from $\bar{\mu}$. Let us first look at the uncertainty's effect on the welfare evaluation of tax reform. Suppose that the uncertainty about θ is small enough such that we can approximate the welfare gain function $W(\mu, \theta)$ by a second-order Taylor expansion at $\bar{\theta}$:

$$W(\mu, \theta) \approx W(\mu, \bar{\theta}) + W_\theta(\mu, \bar{\theta})(\theta - \bar{\theta}) + \frac{1}{2} W_{\theta\theta}(\mu, \bar{\theta})(\theta - \bar{\theta})^2,$$

where W_θ and $W_{\theta\theta}$ represent the first- and second-order partial derivatives of W with respect to θ. We can then write the expected welfare gain from tax reform $\widetilde{W}(\mu)$ as the sum of the welfare gain based on the point estimate $\bar{\theta}$ and an extra term capturing the effect of parameter uncertainty:

$$\widetilde{W}(\mu) = \int_\theta W(\mu,\theta)dF_\theta(\theta) \approx \underbrace{W(\mu,\bar{\theta})}_{\text{based on point estimate}} + \underbrace{\frac{1}{2}\sigma_\theta^2 W_{\theta\theta}(\mu,\bar{\theta})}_{\text{effect of uncertainty}}, \tag{1}$$

where $\sigma_\theta^2 = \int(\theta-\bar{\theta})^2 dF_\theta(\theta)$ is the variance of θ based on policymakers' posterior distribution. Therefore, the existence of parameter uncertainty implies a larger (smaller) welfare gain from the tax reform if $W_{\theta\theta}(\mu,\bar{\theta})$ is positive (negative), i.e., the welfare gain function is strictly convex (concave) in the uncertain parameter.

For the uncertainty's effect on optimal tax policy, from (1), we know that the slope of the expected welfare gain at $\bar{\mu}$ (the optimal tax progressivity based on the point estimate) is given by

$$\widetilde{W}'(\bar{\mu}) \approx \underbrace{W_\mu(\bar{\mu},\bar{\theta})}_{=0} + \frac{1}{2}\sigma_\theta^2 W_{\theta\theta\mu}(\bar{\mu},\bar{\theta}) = \frac{1}{2}\sigma_\theta^2 W_{\theta\theta\mu}(\bar{\mu},\bar{\theta}),$$

where $W_\mu(\bar{\mu},\bar{\theta})$ is the derivative of the welfare gain function with respect to tax progressivity evaluated at $\bar{\mu}$ based on the point estimate $\bar{\theta}$, and it equals zero because $\bar{\mu}$ maximizes $W(\mu,\bar{\theta})$ by definition. Hence the slope of the expected welfare gain function $\widetilde{W}'(\bar{\mu})$ is determined by the third-order derivative $W_{\theta\theta\mu}(\bar{\mu},\bar{\theta})$, which reflects how the convexity of the welfare gain function $W_{\theta\theta}$ varies with tax progressivity μ. If it is positive (negative), the ex ante optimal policy μ^* that maximizes $\widetilde{W}(\mu)$ must be more (less) progressive than $\bar{\mu}$, as long as $\widetilde{W}(\mu)$ is single-peaked.

Proposition 1. *Under previous assumptions, accounting for parameter uncertainty leads to more (less) progressive optimal tax policy if welfare gain from tax reform becomes more (less) convex in the uncertain parameter as tax progressivity increases.*

To better understand this theoretical result, it is helpful to compare the response of optimal tax progressivity to parameter uncertainty here with the precautionary saving behavior of households in the literature since the mechanisms at work are similar. In the context of precautionary savings, households can reduce their utility loss from future consumption risk by increasing current savings in risk-free assets if household prudence is positive, i.e., the third-order derivative of the von Neumann-Morgenstern (vNM) utility function in consumption is positive (Kimball 1990). Notice that precautionary savings do not redistribute consumption across future states, and thus the benefit does not come from a reduction in absolute consumption risk. Instead, precautionary savings increase average future consumption, which shift the distribution of future consumption to an area of the vNM utility function with lower concavity. In other words, the benefit of precautionary savings is through reduction in absolute risk aversion.

Similarly, in the context of optimal policy response to parameter uncertainty, the welfare gain function $W(\mu, \theta)$ is equivalent to the vNM utility function, and its concavity with respect to the uncertain parameter, $-W_{\theta\theta}$, is equivalent to household risk aversion to consumption risk. In the presence of parameter uncertainty, a change in tax policy also cannot transfer resources across different parameter states, and policymakers cannot reduce the absolute parameter risk through tax policy. However, policymakers can still improve the expected welfare by adjusting tax progressivity if it can reduce the "risk aversion", i.e., the concavity of the welfare gain function. Consequently, the desirability of a more or less progressive tax policy in response to parameter uncertainty hinges on the third-order derivative $W_{\theta\theta\mu}$, just like the role of prudence for precautionary savings.

Proposition 1 summarizes the effect of parameter uncertainty when policymakers have limited information about a single parameter. What if there are multiple uncertain parameters? Let $\Theta = (\theta_i)_{i \in I}$ denote the vector of uncertain parameters. Then we can similarly approximate the expected welfare gain from a tax reform by

$$\widetilde{W}(\mu) \approx \underbrace{W(\mu, \bar{\Theta})}_{\text{based on point estimates}} \underbrace{+ \frac{1}{2} \sum_{i,j \in I} W_{\theta_i \theta_j}(\mu, \bar{\Theta}) \sigma_{\theta_i \theta_j}}_{\text{effect of uncertainty}},$$

where $\bar{\Theta}$ is the vector of expected values of parameters, $W_{\theta_i \theta_j}$ is the second-order derivative of W with respect to θ_i and θ_j, and $\sigma_{\theta_i \theta_j}$ is the covariance between θ_i and θ_j implied by the joint posterior distribution of Θ.

The slope of the expected welfare gain at $\bar{\mu}$ is then

$$\widetilde{W}'(\bar{\mu}) \approx \underbrace{W_{\mu}(\bar{\mu}, \bar{\Theta})}_{=0} + \frac{1}{2}\sum_{i,j\in I} W_{\theta_i\theta_j\mu}(\bar{\mu}, \bar{\Theta})\sigma_{\theta_i\theta_j} = \frac{1}{2}\sum_{i,j\in I} W_{\theta_i\theta_j\mu}(\bar{\mu}, \bar{\Theta})\sigma_{\theta_i\theta_j}.$$

Following the previous logic, whether parameter uncertainty leads to more or less progressive optimal tax policy is determined by the sign of $\widetilde{W}'(\bar{\mu})$. However, different from the case with a single uncertain parameter, we can see that now in addition to the shape of the welfare gain function as reflected by the third-order derivatives $\{W_{\theta_i\theta_j\mu}(\bar{\mu}, \bar{\Theta})\}_{i,j\in I}$, the covariances between uncertain parameters $(\sigma_{\theta_i\theta_j})_{i,j\in I}$ implied by the posterior distribution also matter to even the sign of the uncertainty's effect. This suggests that a proper construction of the joint posterior distribution is important for both qualitative and quantitative evaluations about the effects of parameter uncertainty, which is exactly the goal of our Bayesian inference in Section 4.

2.2 Lessons from the Static Model

We now present a static model with standard preferences, for which we can derive the welfare gain function in closed form. We then apply the general conditions from Section 2.1 to determine qualitatively the implications of uncertainty about key economic parameters, namely, the elasticity of labor supply and the magnitude of idiosyncratic wage risk, on progressive income taxation.[8]

Consider a static economy populated by a continuum of measure one households. Households differ in their labor productivity z, which follows a log-normal distribution $LN(-\frac{1}{2}\sigma_z^2, \sigma_z^2)$ in the population. Note that the average productivity is always 1, and σ_z only affects the dispersion of the productivity distribution. Each household chooses labor supply H and consumption C to maximize its utility sub-

[8]Our static model is a simplified version of the model in Heathcote et al. (2017), adding parameter uncertainty. Although the original model of Heathcote et al. (2017) permits a nice closed-form formula for the welfare gain from tax reform without parameter uncertainty, it is no longer the case for the expected welfare gain with parameter uncertainty. Appendix A.3 explores along this direction and illustrate the key difficulty.

ject to the household budget constraint:

$$\max_{\{C,H\}} \ln C - \frac{H^{1+\eta^{-1}}}{1+\eta^{-1}} + \gamma \ln G$$

s.t.

$$C = zH - \widetilde{T}(zH),$$

where η is the Frisch elasticity of labor supply, G is government spending on public consumption, and $\widetilde{T}(\cdot)$ is the income tax function. Following Bénabou (2002) and Heathcote et al. (2017), we set

$$\widetilde{T}(zH) = zH - (1-\chi)(zH)^{1-\mu},$$

where χ controls the tax level, and μ governs the tax progressivity. Let $H(z;\mu,\chi)$ denote the labor supply of productivity z household under policy (μ,χ). The amount of public consumption $G(\mu,\chi)$ is determined by the government budget constraint:

$$G(\mu,\chi) = \int_z \widetilde{T}\left(zH(z;\mu,\chi)\right) dF_z(z),$$

where $F_z(z)$ is the cdf for productivity z.

Policymakers are uncertain about the labor supply elasticity and the amount of idiosyncratic risk, i.e., the uncertain parameters are $\Theta = (\eta, \sigma_z)$. For tractability, we assume that policymakers must choose the tax progressivity μ before parameter uncertainty is resolved, whereas the tax level χ is chosen ex post contingent on the tax progressivity and the true parameter values to maximize the ex post welfare.[9] This simplifying assumption allows us to substitute χ with μ based on the ex post optimality condition, and therefore, we focus on the choice of tax progressivity μ in the following discussion.

Suppose that the status quo income tax is flat, i.e., $\mu_{sq} = 0$,[10] and policymakers are utilitarian. Appendix A.1 shows that the social welfare gain from adopting tax

[9] This assumption will be relaxed in our quantitative analysis.

[10] The status quo tax level is $\chi_{sq} = \frac{\gamma}{1+\gamma}$, which is optimal given $\mu_{sq} = 0$ and independent of parameter Θ.

progressivity μ can be expressed in closed form as

$$\Delta(\mu,\Theta) \equiv \underbrace{\frac{1}{2}\sigma_z^2\mu(2-\mu)}_{\text{insurance benefit}} + \underbrace{\frac{1}{1+\eta^{-1}}\left[(1+\gamma)\ln(1-\mu)+\mu\right]}_{\text{efficiency cost}}.$$

The first part of the formula captures the insurance benefit of progressive income tax against idiosyncratic wage risk, and the second part reflects the efficiency cost of progressive taxation through distorting household labor supply decisions. More specifically, a more progressive income tax discourages labor supply, which reduces the total output ($\ln(1-\mu)$ part) and labor disutility ($+\mu$ part). The insurance benefit is increasing in the amount of idiosyncratic risk σ_z, and the efficiency cost is amplified by labor supply elasticity η. Hence the values of these two parameters directly affect the equity-efficiency trade-off faced by policymakers.

One important issue here is that $\Delta(\mu,\Theta)$ is in household indirect utility, and hence it may not be comparable across different parameter states, especially when there is uncertainty about household preferences. To address this issue, we measure the welfare gain from a tax reform in each parameter state by the consumption equivalent variation, i.e., the amount of consumption transfers required to generate the same welfare gain as the tax reform. Following the convention, we set the consumption transfers to be proportional to household consumption before the tax reform, and given the log utility in consumption, the welfare gain in consumption equivalent variation is

$$W(\mu,\Theta) \equiv \int_z (e^{\Delta(\mu,\Theta)}-1)C_{sq}(z)dF_z(z) = \left(e^{\Delta(\mu,\Theta)}-1\right)\frac{1}{1+\gamma},$$

where $C_{sq}(z) = \frac{z}{1+\gamma}$ is the consumption of productivity z households under the status quo tax policy.

Let $\bar{\Theta} = (\bar{\eta},\bar{\sigma}_z)$ denote the expected values of η and σ_z based on policymakers' posterior belief, and $\bar{\mu}$ denote the policy that maximizes the welfare when uncertainty is ignored, i.e, $W(\mu,\bar{\Theta})$. Appendix A.2 shows that when $\bar{\mu} \in (0,1)$ and $\gamma > 0$, we have $W_{\eta\eta\mu}(\bar{\mu},\bar{\Theta}) > 0$ and $W_{\sigma_z\sigma_z\mu}(\bar{\mu},\bar{\Theta}) > 0$. Applying the results from Section 2.1, and we know that uncertainty about the elasticity of labor supply and the magnitude of idiosyncratic wage risk leads to more progressive optimal income tax.[11]

Intuitively, since the efficiency cost of more progressive income tax is concave in labor supply elasticity, uncertainty about the elasticity lowers the expected efficiency cost following Jensen's inequality, which then implies a more progressive optimal tax policy. Similarly, the insurance benefit of more progressive income tax is convex in the magnitude of wage risk, and hence the expected insurance benefit is higher when the magnitude of wage risk is uncertain, and the optimal tax policy also becomes more progressive.

3 The Quantitative Life-Cycle Model

The sharp theoretical prediction of the static model in Section 2.2 comes at the cost of abstracting away from some aspects of reality important for tax policy consideration, for example, household life cycle and endogenous private insurance through precautionary savings and family labor supply. To provide a quantitative examination of uncertainty's effect on tax policy, a richer model augmented with these additional features is called for. Our analysis will focus on policymakers' uncertainty about household preferences and idiosyncratic wage process because there is, arguably, more consensus on model specification, and parameter uncertainty is of major concern for tax policy decisions.[12]

In this section, we describe the quantitative dynamic model, through which we interpret the data and measure uncertainty about preference and wage parameters. The model also serves as a laboratory for conducting thought experiments of tax reforms and evaluating their welfare consequences. We first describe the physical

[11] For simplicity, we assume here that there is no correlation between the value of η and σ_z based on policymakers' belief, i.e., $\sigma_{\eta\sigma_z} = 0$. We will allow correlations between uncertain parameters in our quantitative analysis.

[12] In contrast, we do not model explicitly certain factors (e.g. endogenous skill accumulation) in our quantitative model because i) there is greater uncertainty about the mechanism (e.g., learning-by-doing or on-the-job-training) than uncertainty about related parameters, and ii) there are often identification issues without proper data (e.g., Heckman et al. 2002). Hence we take a more agnostic approach and try to account for their effects on tax policy later through the Pareto weights in the optimal tax policy problem (details in Section 5). Also, taking into account parameter uncertainty in the policymaking problem requires numerical integration with respect to the multivariate posterior distribution of parameters, which increases the computation burden significantly compared to the standard optimal tax policy exercises in the literature, and thus keeping the model to the point also help ease this tension.

environment of our model, and then state the household optimization problem in recursive formulation.

3.1 Environment

We study a partial equilibrium life-cycle model with idiosyncratic wage risk and endogenous labor supply. We follow a cohort of a continuum of measure one households over their life cycle. These households live for T periods, from age 1 to T, work in the first R periods of life, and then are retired from age $R+1$ onward. Each household consists of two members: a male and a female. For simplicity, we omit the index for different households and denote by $X_{j,t}$ the variable X of member j at age t, with $j = 1$ or 2 indicating the male or the female member.

Households enjoy joint consumption C_t and choose the labor supply of both members $H_{1,t}$ and $H_{2,t}$ that incur disutility. The period utility function is given by $u(C_t, H_{1,t}, H_{2,t})$. An operative extensive margin of female labor supply is included in the model by introducing a fixed per-period utility cost f whenever female hours worked are strictly positive. Households discount the future utility at the constant rate δ, so that $1/(1+\delta)$ is the household time discount factor.

Members of a household can work at wages $W_{j,t}$ determined by their labor productivity. Log-wages of both household members are stochastic and represent the sum of i) a deterministic life-cycle component $g_{j,t}$ that is common across households, and ii) an idiosyncratic stochastic component $F_{j,t}$:

$$\ln W_{j,t} = g_{j,t} + F_{j,t}. \tag{2}$$

The idiosyncratic component $F_{j,t}$ follows an AR(1) process with persistence ρ_j:

$$F_{j,t} = \rho_j F_{j,t-1} + v_{j,t},$$

where $v_{j,t}$ is the normally distributed random shock to member j's wage at age t, and they may be correlated between two members of the same household, but are independent over time and across different households:

$$\begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma^2_{v_1} & \sigma_{v_1,v_2} \\ \sigma_{v_1,v_2} & \sigma^2_{v_2} \end{bmatrix}\right).$$

Here $\sigma^2_{v_1}$, $\sigma^2_{v_2}$, and σ_{v_1,v_2} are the variances and covariance of male and female wage shocks. Note that both the deterministic component and the parameters governing the stochastic component are gender-specific.

As is common in standard incomplete-market models, households cannot trade fully state-contingent Arrow securities, but they can save, and potentially borrow, at the risk-free interest rate r subject to a borrowing limit $\underline{A}$.[13] Working-age households need to pay income and payroll taxes in each period, and retired households are eligible for a fixed amount of retirement benefit b in each period in which she is alive.[14]

3.2 Household Optimization Problem

The state variables of a working-age household include the current savings A, the male and female idiosyncratic wage components, F_1 and F_2, and the age of the household t. In each period, members of each household make joint decisions on household consumption C, savings for the next period A', and labor supply of both members H_1 and H_2 to maximize their discounted utility subject to the budget constraint. A working-age household's problem is then, in recursive form:

$$\begin{aligned} V(A, F_1, F_2, t) = & \max_{\{C, A', H_1, H_2\}} u(C, H_1, H_2) - \mathbf{I}(H_2 > 0) f \\ & + \frac{1}{1+\delta} \mathbb{E}_{(F_1', F_2')} \left[V(A', F_1', F_2', t+1) | F_1, F_2 \right] \\ \text{s.t.} \quad & C + A' = Y - \widetilde{T}(Y) - \tau_{ss} Y + (1+r)A, \\ & Y = W_{1,t} H_1 + W_{2,t} H_2, \end{aligned}$$

[13]Households are born with zero savings.

[14]The U.S. social security benefits are piecewise linear functions of average monthly past earnings over the working life. Additional rules govern benefits for spouses. A full representation of the U.S. social security system is costly in terms of computation, since it adds two continuous state variables to the recursive formulation of the problem. Hence we model the U.S. social security benefit formula starkly, by assuming that the benefits per household are independent of past contributions.

$$C, H_1, H_2 \geq 0,\ A' \geq \underline{A},$$

where $\mathbf{I}(H_2 > 0)$ equals 1 if female hours H_2 is positive. Female hours of $H_2 = 0$ corresponds to non-participation. The term $\widetilde{T}(Y)$ in the budget constraint is the income tax function that determines the tax liability of a household with before-tax income Y, and τ_{ss} is a flat payroll tax representing the Federal Insurance Contribution Act (FICA) taxes. The wage of each member $W_{j,t}$ is determined by household states F_j and t according to (2).

After retirement, labor productivity falls to zero, and hence households optimally do not work in retirement. The state variables of retired households reduce to only current savings and the age of the household. The dynamic programming problem of a retired household is then:

$$\begin{aligned} V^R(A, t) = \max_{\{C, A'\}} & \ u(C, 0, 0) + \frac{1}{1+\delta} V^R(A', t+1) \\ \text{s.t.} \quad & C + A' = b + (1+r)A, \\ & C \geq 0,\ A' \geq \underline{A}. \end{aligned}$$

Households are assumed to have an additively separable utility function of the form:

$$u(C, H_1, H_2) = \frac{C^{1-\sigma}}{1-\sigma} - \psi_1 \frac{H_1^{1+\eta_1^{-1}}}{1+\eta_1^{-1}} - \psi_2 \frac{H_2^{1+\eta_2^{-1}}}{1+\eta_2^{-1}}.$$

The advantage of using this preference structure is that all the preference parameters are directly interpretable. The parameter σ governs household risk aversion, and its reciprocal is the Frisch elasticity of consumption with respect to its own price. The parameters η_1 and η_2 are the Frisch elasticities of male and female labor supply with respect to their own wages, and ψ_1 and ψ_2 control the levels of disutility from male and female labor supply.

We permit the income tax function $\widetilde{T}(Y)$ to be progressive and use the same two-parameter tax function as in the static model:

$$\widetilde{T}(Y) = Y - (1-\chi) Y^{1-\mu}, \tag{3}$$

where μ and χ are two parameters governing the progressivity and the level of the income tax, respectively. It implies that after-tax income $Y - \widetilde{T}(Y)$ is an increasing and concave function of pre-tax income Y.

4 Measure Uncertainty from the Data

In this section, we take the model in Section 3 to the U.S. data and quantify the degree of uncertainty about parameters governing household preferences and idiosyncratic wage process. We first describe the main data set used for our empirical analysis and how we choose the values of externally calibrated parameters. We then provide a brief summary of the limited-information Bayesian method that we employ to estimate the preference and wage parameters. After that, we explain our estimation strategy and report the estimation results. Finally, we assess the goodness of fit of our model.

4.1 The Data

Our main data source is the core sample of the 1999-2017 Panel Study of Income Dynamics (PSID), from which we obtain individual and household level information about earnings, hours worked, consumption, net worth, and characteristics such as age, race, education, number of children, and so on.[15] The longitudinal structure of the PSID allows us to follow the same individuals and households over time (i.e., every two years) and observe the comovement in these variables, which is crucial for our estimation strategy. We focus on married households with working male head aged between 30 and 57 because this group of households fit best the specification of our life-cycle model and represent the majority of the U.S. population. To ease the burden of computation, we set each period in our life-cycle model to four years in the data, and hence the biennial PSID data are converted to four-year frequency. Wages are constructed as earnings divided by hours worked.

All nominal variables are converted to values in year 2000 U.S. dollars based on the consumer price index for all urban consumers (CPI-U) from the U.S. Bureau of Labor Statistics. We set the units of income and labor supply in our model to the

[15] We follow closely the procedures in Blundell et al. (2016) when constructing variables from the PSID data and identifying outliers.

average four-year earnings and hours worked by males in our PSID sample, which are \$243,766 and 9,202 hours (i.e., \$60,942 and 2301 hours per year), respectively.

4.2 External Calibration

Before estimating the preference and wage parameters, we first calibrate the other parameters outside of the life-cycle model, and their values are reported in Table 1. The key difference between these calibrated parameters and the estimated ones is that the former are either directly observable or can be inferred with few to no model assumptions.

4.2.1 Demographic

As mentioned earlier, one period in the model represents four years in the data. We consider the part of household life cycle between age 22 and age 81 with a retirement age of 65. This then translates to a life cycle in the model from age 1 to $T = 15$ with a retirement age of $R = 11$.

4.2.2 Income Tax

There are two parameters in the income tax function given by (3): μ for the income tax progressivity and χ for the income tax level. We estimate these two parameters by taking the natural log of (3) and running the following OLS regression with our PSID sample:

$$\ln(Y - \widetilde{T}(Y)) = \ln(1 - \chi) + (1 - \mu) \ln(Y).$$

Tax liability $\widetilde{T}(Y)$ is defined as federal income tax minus earned income tax credit (EITC) and food stamp benefits. Federal income tax and EITC are calculated based on household income Y and the actual income tax code, and food stamp benefits are obtained directly from the PSID. The estimated income tax parameters are $\mu = 0.128$ and $\chi = 0.126$, with standard error 0.002 and 0.001, respectively.

4.2.3 Payroll Tax and Retirement Benefit

The flat payroll tax rate τ_{ss} is set to 7.65%, based on the Federal Insurance Contribution Act (FICA) tax rates on pre-tax income of employees. The retirement benefit b in the model corresponds to the sum of Social Security benefits and benefits from Medicare. We calibrate the Social Security benefits to the sum of average benefits received by males and females aged 66 and older in the 1999-2017 Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS), given by \$20,975 per year in 2000 dollars. Since the benefits from Medicare are difficult to measure directly, we assume that they are proportional to the social security benefits, based on the ratio of Medicare tax rate to Social Security tax rate. Therefore, the retirement benefit b in the model is calibrated to $\$20,975 \times 7.65\%/6.2\% \times 4 = \$103,522$, i.e., 0.425 in model income units.

4.2.4 Wage Trend and Initial Wage

For the deterministic life-cycle component of log-wage, $g_{j,t}$ in (2), we regress the male or female log-wage on a quadratic polynomial in age, together with a group of controls for the year, education, race, and location, etc. The gender-specific age-profiles of log-wage are then constructed as the predicted values from these regressions at different ages while integrating over the remaining covariates. The resulting wage trends are presented in Figure 9 of Appendix B. For the initial idiosyncratic component of log-wage at the labor market entry, $F_{j,1}$ in (2), we assume that it is normally distributed across households and potentially correlated between the two members of each household. We estimate the covariance matrix of the initial idiosyncratic components from the residuals of the log-wage regressions at model age 1 (data age 22-25). The variances of male and female initial wages are 0.211 and 0.232, respectively, and the initial wages are positively correlated with a correlation coefficient of 0.214.

4.2.5 Interest Rate and Borrowing Limit

We set the annual risk-free interest rate at 2%, which implies a four-year interest rate of $r = 8.24\%$. We exclude non-collateralized debts in the model by imposing a zero borrowing limit (i.e., $\underline{A} = 0$).[16]

[16] Since we focus on households aged 30-57, who often have significantly positive net worth, the zero-borrowing-limit assumption is inconsequential to our results.

Table 1: Externally Calibrated Parameters

Parameter	Description	Value
T	length of life cycle	15
R	retirement age	11
μ	income tax progressivity	0.128
χ	income tax level	0.126
τ_{ss}	payroll tax rate	7.65%
b	retirement benefit	0.425
$\sigma^2_{F_{1,1}}$	variance of male initial wage	0.211
$\sigma^2_{F_{2,1}}$	variance of female initial wage	0.232
$corr_{F_{1,1},F_{2,1}}$	correlation of male and female initial wage	0.214
r	real interest rate	8.24%
$\underline{A}$	borrowing limit	0

Notes: Each period in the model corresponds to four years in the data.

4.3 Bayesian Inference

We estimate the remaining parameters about household preferences and idiosyncratic wage process using the PSID data through a limited-information Bayesian approach and quantify the uncertainty about these parameters by the posterior distribution.

4.3.1 The Limited-Information Bayesian Method

The limited-information Bayesian method, as described in Kim (2002) and later advocated by Christiano et al. (2010) and Fernández-Villaverde et al. (2016) among others, can be viewed as the Bayesian version of the generalized method of moments (GMM). Similar to GMM, the limited-information Bayesian method only uses a set of moments from the data for parameter inference, and therefore, it does not require strong distributional assumptions about the error terms in the model of the data generating process. There are two main reasons why we adopt the limited-information Bayesian method for our empirical analysis. First, a proper characterization of parameter uncertainty requires the distribution of parameters conditional on the information we already have (e.g., the data we observe), which is conveniently the output of Bayesian inference.[17] Second, the full likelihood function is

[17] In contrast, the standard GMM produces estimators for parameters as functions of the data with distributions conditional on the true values of parameters.

difficult to construct for a complex nonlinear model like ours, whereas the limited-information Bayesian method allows us to sidestep this obstacle.

Let Θ denote the parameters of interest and $\hat{\mathbf{m}}$ denote the vector of M empirical moments from the data for estimation. Kim (2002) shows that the likelihood of $\hat{\mathbf{m}}$ conditional on Θ is approximately

$$f(\hat{\mathbf{m}}|\Theta) = (2\pi)^{-\frac{M}{2}} |S|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\hat{\mathbf{m}} - \mathbf{m}(\Theta)\right)' S^{-1} \left(\hat{\mathbf{m}} - \mathbf{m}(\Theta)\right)\right], \tag{4}$$

where $\mathbf{m}(\Theta)$ is the model's prediction for the moments under parameter Θ, and S is the covariance matrix of $\hat{\mathbf{m}}$. The covariance matrix S is often unknown but can be replaced by a consistent estimator of it, which can be obtained through bootstrap. Bayes' theorem tells us that the posterior density $f(\Theta|\hat{\mathbf{m}})$ is proportional to the product of the likelihood $f(\hat{\mathbf{m}}|\Theta)$ and the prior density $p(\Theta)$:

$$f(\Theta|\hat{\mathbf{m}}) \propto f(\hat{\mathbf{m}}|\Theta)p(\Theta), \tag{5}$$

and we can then apply the standard Markov Chain Monte Carlo (MCMC) techniques such as the Metropolis-Hastings algorithm to obtain a sequence of random samples from the posterior distribution.

4.3.2 Estimation Strategy

Except the Bayesian approach, we follow closely the empirical strategy in Blundell et al. (2016). In particular, using the PSID data, we first regress wage, consumption, and earnings growth on observable characteristics of individuals and households and obtain the residuals $\Delta w_{j,t}$, Δc_t, and $\Delta y_{j,t}$, respectively. Similar to Blundell et al. (2016), we estimate the five wage parameters (ρ_1, ρ_2, $\sigma^2_{v_1}$, $\sigma^2_{v_2}$, $corr_{v_1,v_2}$) using only a group of second-order moments of $\Delta w_{j,t}$, and hence the estimation of wage parameters requires only the statistical model of wage process and is immune to our assumptions about the life-cycle model.[18] Given the estimates of wage parameters, we then combine a group of second-order moments of $\Delta w_{j,t}$, Δc_t, and $\Delta y_{j,t}$ and a group of first-order moments of earnings, hours worked, female partici-

[18] We estimate the correlation coefficient between male and female wage shocks $corr_{v_1,v_2}$ instead of the covariance σ_{v_1,v_2}.

pation, and household net worth to estimate jointly the seven preference parameters $(\sigma, \eta_1, \eta_2, \psi_1, \psi_2, f, \delta)$.[19] Blundell et al. (2016) show that the second-order moments identify the elasticity parameters (σ, η_1, η_2), and the remaining preference parameters are pinned down by the first-order moments.[20] The list of moment conditions that we employ in our estimation are provided in Table 8 of Appendix B.

When constructing the likelihood in (4) for Bayesian estimation, the estimate of the covariance matrix S is obtained from 5,000 bootstrap samples of the PSID data, and the model predicted moments $\mathbf{m}(\Theta)$ are computed based on a model-simulated panel of 50,000 households.[21] We adopt an uninformative prior $p(\Theta)$ consisting of independent uniform distributions of each parameter. We apply the random-walk Metropolis-Hastings algorithm to simulate draws from the posterior density $f(\Theta|\hat{\mathbf{m}})$ given by (5), and the posterior distribution is characterized by a sequence of 15,000 draws after a burn-in of 5,000 draws.

4.3.3 Estimation Results

Table 2 reports the posterior means and the 95% credible intervals of preference and wage parameters from the Bayesian estimation, together with the uniform priors. For the posterior means, one interesting result is that our estimate of the Frisch elasticity of female labor supply η_2 is somewhat lower than the male counterpart η_1. The reason is partly because we allow gender-specific persistence of wage shocks and find that female shocks are significantly less persistent than male shocks (0.762 vs. 0.917). This difference reduces the income/wealth effect of female wage shocks, and hence a weaker substitution effect (i.e., a lower elasticity of labor supply) for females is enough to explain the variations of female earnings in the data. The 95% credible intervals of the posterior distributions are much narrower than the uniform priors, suggesting that the uncertainty about these parameters is greatly reduced by

[19]Note that we do not need to adjust the empirical moments for sample selection due to endogenous female participation decisions since our structural model also features an operative extensive margin of female labor supply that is disciplined by the data.

[20]We verify our empirical strategy with model-simulated data and find that it works well in recovering the true values of parameters.

[21]For consistency, we add measurement errors to model-simulated data as well. Like Blundell et al. (2016), the standard deviation of measurement errors on log-consumption is set to 0.20 based on the data moment $\mathbb{E}(\Delta c_t \Delta c_{t-1})$. For log-earnings and log-hours, the standard deviations of measurement errors are set to 0.15 following the literature.

the information contained in the data.

Table 2: Estimated Parameters

Parameter	Description	Posterior Distribution Mean	Posterior Distribution 95% Interval	Uniform Prior [Min, Max]
σ	household risk aversion	2.43	[1.95, 2.86]	[1.00, 6.00]
η_1	male labor supply elasticity	0.335	[0.257, 0.420]	[0.001, 2.000]
η_2	female labor supply elasticity	0.234	[0.182, 0.292]	[0.001, 2.000]
ψ_1	disutility of male labor supply	0.901	[0.808, 0.995]	[0.010, 5.000]
ψ_2	disutility of female labor supply	3.23	[2.22, 4.67]	[0.01, 5.00]
f	fixed utility cost of female participation	0.159	[0.142, 0.176]	[0.001, 0.500]
δ	discount rate of utility	0.0273	[0.0201, 0.0346]	[-0.1000, 0.1000]
ρ_1	persistence of male wage shocks	0.917	[0.785, 1.012]	[0.100, 1.200]
ρ_2	persistence of female wage shocks	0.762	[0.597, 0.931]	[0.100, 1.200]
$\sigma^2_{v_1}$	variance of male wage shocks	0.0736	[0.0643, 0.0815]	[0.0001, 0.3000]
$\sigma^2_{v_2}$	variance of female wage shocks	0.0844	[0.0759, 0.0925]	[0.0001, 0.3000]
$corr_{v_1,v_2}$	correlation of male and female wage shocks	0.064	[0.013,0.114]	[-0.300, 0.500]

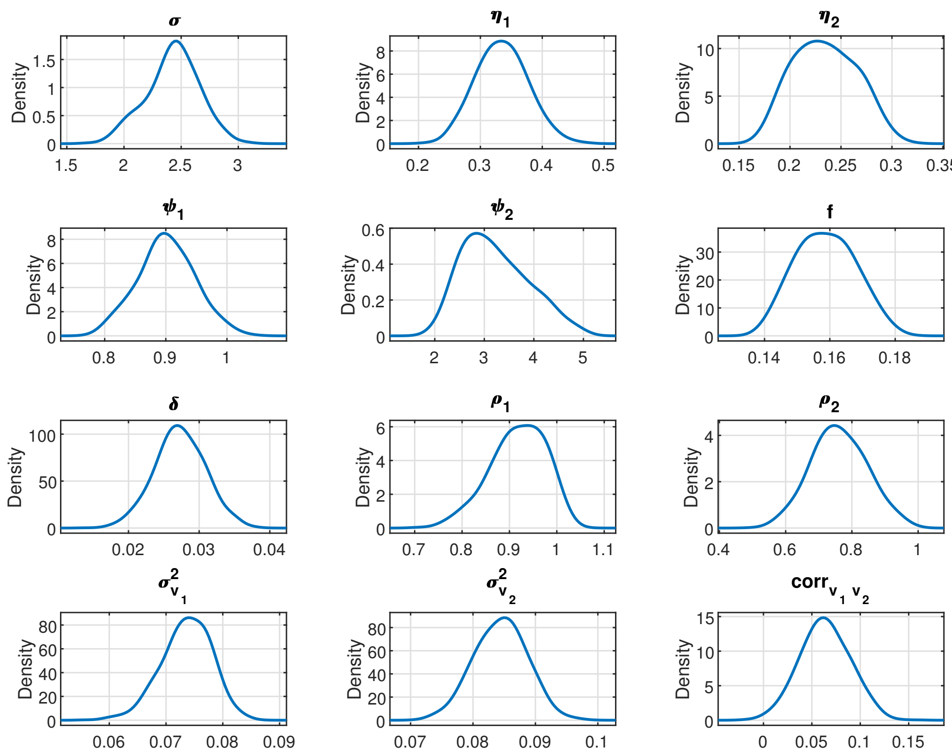

Figure 1: Distributions of Preference and Wage Parameters

Notes: This figure plots the marginal probability density functions of preference and wage parameters implied by the joint posterior distribution.

Figure 1 plots the posterior distributions of each preference and wage parameter. Notice that the posterior distributions are not necessarily normal or symmetric. Both Table 2 and Figure 1 are about the marginal distributions of parameters, but since these parameters are estimated jointly, they may be correlated in the posterior distribution. Table 9 and 10 of Appendix B report the correlations among preference and wage parameters, respectively. In general, the correlations among preference parameters are stronger than those among the wage parameters.

4.4 Goodness of Model Fit

How well does the model fit the data? Table 3 reports a group of first-order moments of household variables from the data, along with their counterparts implied by the model. The moments include average net worth of households, average male and female earnings and hours worked, and female non-participation rate. It is worth mentioning that the first-order moments are employed jointly with the second-order moments to estimate the preference parameters, and we have a heavily overidentified system. Therefore, the first-order moments are not matched mechanically. Nevertheless, for all moments the model fits the data well.

Table 3: Data vs. Model: First Moments

Moment	Description	Data	Model
$\mathbb{E}[A_t]$	average net worth	0.884	0.902
$\mathbb{E}[Y_{1,t}]$	average male earnings	1	0.941
$\mathbb{E}[Y_{2,t} \mid H_{2,t} > 0]$	average female earnings | work	0.481	0.468
$\mathbb{E}[H_{1,t}]$	average male hours	1	1.002
$\mathbb{E}[H_{2,t} \mid H_{2,t} > 0]$	average female hours | work	0.706	0.726
$\mathbb{E}[\mathbf{I}(H_{2,t} = 0)]$	female non-participation rate	0.154	0.157

Notes: "Model" results are the mean of each moment implied by the model and the posterior distribution of parameters.

Figure 2 plots the life-cycle profiles of household endogenous variables, namely household consumption, asset, and hours worked, for both the model (blue solid lines) and the data (red dotted lines with shaded 95% confidence intervals). Although not targeted in calibration or estimation, the model still matches these life-cycle profiles reasonably well. Consumption in the model rises over the life cycle since wage risk, and the associated precautionary saving, as well as a fairly high degree of patience lead to low consumption early in life and subsequent positive consumption growth. Asset rises over the life cycle as households accumulate wealth to fund retirement consumption, but also for precautionary reasons to hedge against stochastic wage fluctuations. Lower average female hours originate both from lower hours conditional on working, but also from a significant non-participation rate. The model captures well the growth of consumption and asset over the life cycle in the data, as well as the stable male and female hours.

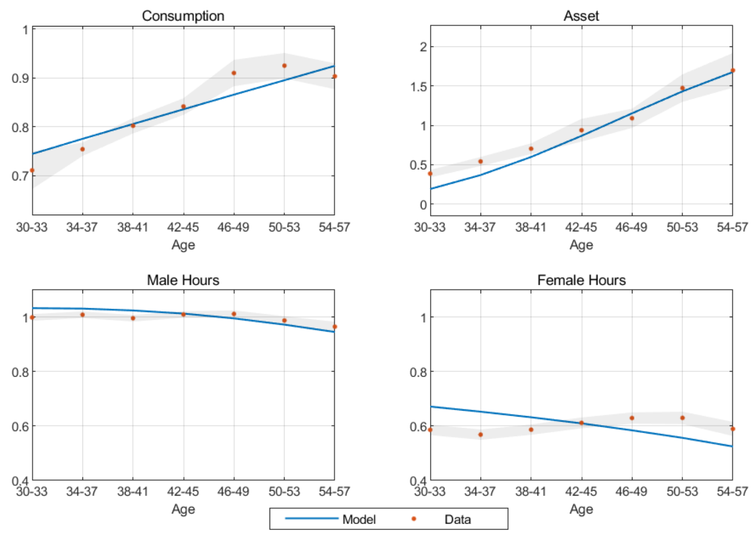

Figure 2: Data vs. Model: Life Cycles

Notes: This figure shows the life cycles of cross-sectional means in the model with parameters equal to the posterior means (blue solid lines) and in the data (red dotted lines) together with the 95% confidence interval (grey bands). Since the data do not include all types of consumption expenditures, the consumption life cycle from the data is scaled up to match the average consumption in the model. For ease of comparison, male and female hours are placed on the same scale.

Figure 3 shows the performance of the model in matching the second-order moments of wage, consumption, and earnings growth in the data, which are important for inequality. The model does an excellent job in predicting moments of the joint distribution of male and female wage growth (the left panel). This suggests that the AR(1) model provides a good approximation to the actual wage process in the data.

The model also matches well the volatility of consumption and earnings growth (top middle panel). The minor discrepancies in the variances of consumption and female earnings growth are driven by the conflict between their variances and co-

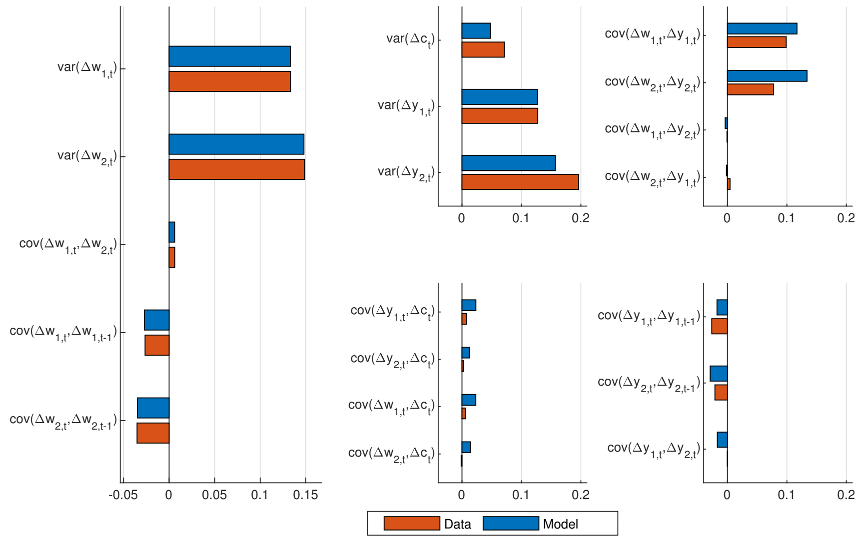

Figure 3: Data vs. Model: Second Moments

Notes: This figure shows the second moments of wage, consumption, and earnings growth in the model and in the data. "Model" results are the mean of each moment implied by the model and the posterior distribution of parameters. For ease of comparison, the middle and right panels are plotted on the same scale.

variances in the data. Specifically, the covariances between consumption and earnings/wages are fairly small in the data (bottom middle panel), which suggests that consumption is not affected much by wage shocks and therefore should not be too volatile. This, however, contradicts the large variance of consumption growth in the top middle panel. Since the model tries to match these moments simultaneously, it undershoots the variance and overpredicts the covariances. A similar case can be made for female earnings (top right panel vs. top middle panel)

Finally, in the bottom right panel we plot lagged auto- and cross-covariances of male and female earnings. Consistent with the data, male and female earnings growth in the model have negative autocorrelations, and their contemporaneous covariance is close to zero.

Overall, the evidences above suggest that our life-cycle model fits the data well, albeit not perfectly, and hence it is suitable for a quantitative examination of parameter uncertainty's effects on income tax policy.

5 Tax Policy Implications of Uncertainty

In this section, we study quantitatively the income tax policy implications of policymakers' uncertainty about household preferences and wage process, based on the life-cycle model in Section 3 and the empirical posterior distribution in Section 4. We first introduce the optimal tax policy problem with parameter uncertainty and investigate how the uncertainty about household preferences and wage process affects the optimal design of income tax code. We then evaluate the welfare cost of this uncertainty by comparing the ex ante optimum to the first-best scenario with perfect information. Finally, we explore implications of various risk preferences of policymakers.[22]

5.1 Uncertainty and Optimal Tax Policy

5.1.1 The Optimal Tax Problem

We now describe the optimal tax policy problem of policymakers who are uncertain about household preferences and wage process. Briefly speaking, policymakers face the classic equity-efficiency trade-off when choosing the tax policy. On the one hand, a progressive income tax provides valuable public insurance against idiosyncratic wage risks and reduces inequality. On the other hand, a progressive income tax leads to efficiency costs through distortions to household labor supply. Both the benefits and costs of progressive income tax are regulated by the underlying parameters governing household preferences and wage process, and so is the optimal tax policy.[23] However, when policymakers are uncertain about the values of these key parameters, there is a new dimension of policy consideration: that is, the policymakers must also balance the performance of tax policy and trade off welfare gains and losses across possible scenarios with different parameter values.

Specifically, tax policy is represented by the two parameters of the income tax function in (3), μ and χ, which control the income tax progressivity and level, re-

[22] All quantitative results are based on model-simulated panels of 50,000 households over their life cycles. More details about computation method are in Appendix B.3.

[23] For example, the value of public insurance depends on the variance and persistence of wage shocks, household risk aversion, and the extent of self-insurance through precautionary savings and family labor supply; and the efficiency costs are determined by the elasticity of labor supply, etc.

spectively. Policymakers' uncertainty is summarized by a posterior distribution $\Pi(\Theta)$, where Θ is the vector of model parameters governing household preferences and wage process. Let $W(\mu, \chi, \Theta)$ denote the objective function of policymakers if they know Θ for sure, the (ex ante) optimal tax policy (μ^*, χ^*) is then given by

$$(\mu^*, \chi^*) = \arg\max_{(\mu,\chi)} \Gamma^{-1}\left(\int_{\Theta} \Gamma\left(W(\mu, \chi, \Theta)\right) d\Pi(\Theta)\right), \tag{6}$$

where $\Gamma(\cdot)$ is a strictly increasing function reflecting the risk preferences of policymakers. The inverse function $\Gamma^{-1}(\cdot)$ outside converts the policymakers' "utility" to its certainty equivalent value. For our benchmark analysis, policymakers are assumed to be risk-neutral, i.e., $\Gamma(\cdot)$ is linear. We investigate the implications of policymakers' risk aversion in Section 5.3.

Given parameter Θ, policymakers' goal is simply to maximize the welfare gain from adopting the new policy (μ, χ), i.e., $W(\mu, \chi, \Theta)$. Since policymakers are uncertain about the true value of Θ, when choosing the tax policy, they must trade off welfare gains and losses across parameter states in which household preferences differ. Hence to make such comparison meaningful, we measure $W(\mu, \chi, \Theta)$ in consumption equivalent variations. More specifically, suppose that the social welfare function $\mathrm{SWF}(\mu, \chi, \Theta)$ is given by a weighted sum of household welfare for a newborn cohort:[24]

$$\mathrm{SWF}(\mu, \chi, \Theta) = \int_{\mathbf{s}} \omega(\mathbf{s}) \widetilde{V}_1(\mathbf{s}; \mu, \chi, \Theta) d\Phi_1(\mathbf{s}),$$

where $\mathbf{s} = \{A, F_1, F_2\}$ is the vector of household states except age, $\omega(\mathbf{s})$ is the Pareto weight function, $\widetilde{V}_1(\mathbf{s}; \mu, \chi, \Theta)$ is the expected lifetime utility of a newborn (i.e., age 1) household, and $\Phi_1(\mathbf{s})$ is the distribution of newborn households, which is independent of tax policy and uncertain parameters. $W(\mu, \chi, \Theta)$ is then defined as the total amount of consumption transfers required to achieve the same social welfare change as adopting tax policy (μ, χ), i.e., $\mathrm{SWF}(\mu, \chi, \Theta) - \mathrm{SWF}(\mu_{sq}, \chi_{sq}, \Theta)$, where (μ_{sq}, χ_{sq}) denotes the status quo tax policy (more details in Appendix B.2).

Households derive utility from both private and public consumption, and the

[24]We choose the welfare measure for a newborn cohort to avoid the redistribution motive of tax policy over different cohorts and hence the need to take a stand on how to weight different generations.

expected lifetime utility $\widetilde{V}_1(\mathbf{s};\mu,\chi,\Theta)$ is given by:

$$\widetilde{V}_1(\mathbf{s};\mu,\chi,\Theta) \equiv V_1(\mathbf{s};\mu,\chi,\Theta) + \gamma\frac{[G(\mu,\chi,\Theta)]^{1-\sigma_G}-1}{1-\sigma_G}.$$

Here $V_1(\mathbf{s};\mu,\chi,\Theta)$ is the value function of a newborn household defined by the household optimization problem in Section 3.2, which combines utility from private consumption and labor disutility. $G(\mu,\chi,\Theta)$ denotes the amount of public consumption, which is financed by tax revenues from the same cohort:

$$G(\mu,\chi,\Theta) = \sum_{t=1}^{R}\frac{\mathrm{Tax}_t(\mu,\chi,\Theta)}{(1+r)^{t-1}},$$
$$\mathrm{Tax}_t(\mu,\chi,\Theta) = \int_{\mathbf{s}}\widetilde{T}(Y_t(\mathbf{s};\mu,\chi,\Theta))d\Phi_t(\mathbf{s};\mu,\chi,\Theta),\ t=1,\ldots,R,$$

where $Y_t(\mathbf{s};\mu,\chi,\Theta)$ is household income, $\Phi_t(\mathbf{s};\mu,\chi,\Theta)$ is the distribution of households, and $\mathrm{Tax}_t(\mu,\chi,\Theta)$ is the tax revenue, all for age-t households. Public consumption benefits all households equally, and γ is the parameter that controls the taste for public consumption relative to private consumption, and σ_G governs how fast the marginal utility of public consumption declines with its level. We set σ_G to the posterior mean of its counterpart for private consumption σ.

5.1.2 Uncertainty's Effect on Optimal Tax Policy

To quantify the effects of uncertainty about household preferences and wage process on the optimal design of tax policy, we solve numerically two versions of the optimal tax policy problem described in Section 5.1.1. In one version, policymakers ignore the uncertainty and treat the point estimates (i.e., posterior means) of parameters $\bar{\Theta}$ as their true values when searching for the optimal tax policy. This is equivalent to using a degenerate posterior distribution with $\Pr(\Theta=\bar{\Theta})=1$ in the optimal tax policy problem. In the other version, policymakers take into account the uncertainty correctly in the policymaking process by using the true posterior distribution estimated from the data in Section 4.3. The difference in optimal tax policy between the two versions then reveals how tax policy should adjust in response to the existence of such uncertainty.

To ensure that our quantitative findings are pertinent to practical policy considerations, we calibrate the Pareto weights $\omega(\mathbf{s})$ in the social welfare function and the taste for public consumption γ such that the status quo policy is optimal based on the point estimates of parameters. Therefore, policymakers would have no incentive to deviate from the status quo policy except their concerns about parameter uncertainty. Such calibration strategy also serves as a reduced-form way of controlling for the effects of other economic and non-economic determinants of tax policy that are not modeled explicitly in our framework. Hence our quantitative analysis shall be understood as investigating the tax policy implications of uncertainty about household preferences and wage process, while holding constant the effects of other factors.

In particular, similar to Chang et al. (2018) and Heathcote and Tsujiyama (2021) , the Pareto weight function takes the following functional form:

$$\omega(\mathbf{s}) = \frac{\exp\left[\xi\left(W_{1,1}(\mathbf{s}) + W_{2,1}(\mathbf{s})\right)\right]}{\int_{\mathbf{s}} \exp\left[\xi\left(W_{1,1}(\mathbf{s}) + W_{2,1}(\mathbf{s})\right)\right] d\Phi_1(\mathbf{s})},$$

where $W_{1,1}(\mathbf{s})$ and $W_{2,1}(\mathbf{s})$ are the male and female wage of the newborn household, and ξ is a parameter that controls policymakers' "taste for redistribution": when $\xi = 0$, policymakers are utilitarian; a larger ξ implies that the policymakers value the welfare of high-wage households more, and thus a less progressive tax system is preferred. One the other hand, a larger γ indicates greater benefits from public consumption, which leads to a higher level of optimal income tax. The values of ξ and γ that rationalize the status quo policy are 1.59 and 4.17, respectively.

Table 4: Optimal Tax Policy

	Progressivity (μ)	Level (χ)
Based on Point Estimates	0.128	0.126
Based on Posterior Distribution	0.147	0.120

Notes: By construction, the optimal policy based on point estimates are the same as the status quo policy.

The optimal tax policies based on the point estimates and the true posterior distribution are reported in Table 4. By construction, the optimal policy based on the point estimates are exactly the same as the status quo policy reported in Table 1.

Taking into account the uncertainty about preference and wage parameters as represented by the posterior distribution, however, leads to a more progressive optimal tax policy, with an increase in progressivity by 2.0 percentage points (0.147 vs. 0.128). The level of the optimal tax policy also falls by 0.6 percentage point (0.120 vs. 0.126). To get a better sense of what these differences imply about the income tax rates, Figure 4 plots the changes in marginal tax rates faced by households of various income levels. For households near the top of the income distribution (the right limit of the horizontal axes), their marginal tax rates would rise by about 3.0%. For households near the bottom of the income distribution (the left limit of the horizontal axes), their marginal tax rates would fall by about 2.2%. Hence, accounting for the uncertainty would raise the gap in marginal tax rate between the high- and low-income households by about 5.2%.

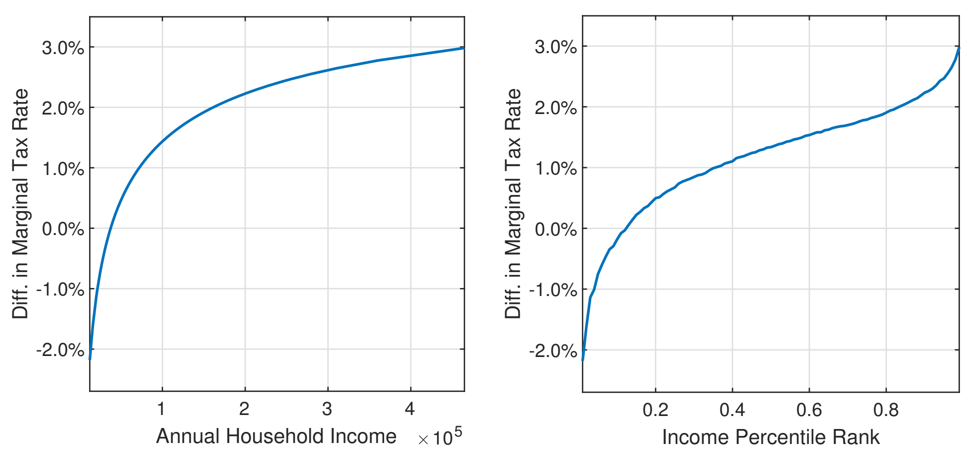

Figure 4: Uncertainty's Effect on Marginal Tax Rate

Notes: This figure plots the change in the optimal income tax scheculde (i.e., marginal tax rate) after accounting for the uncertainty about preference and wage parameters represented by the posterior distribution. The left panel is against household income in 2017 dollars, and the right panel is against the percentile rank of household income. The lower and upper limits of the horizontal axes correspond to the bottom and top 1% cutoff levels of the income distribution for married households.

5.1.3 Why More Progressive Optimal Tax?

Why is the optimal tax policy more progressive with the uncertainty about preference and wage parameters? The answer is two-fold.[25] First, why does the uncer-

[25]Our discussion here focus on the ideas and intuitions behind the quantitative findings, which echoes our formal theoretical analysis in Section 2.

tainty matter at all for the choice of tax policy? Obviously, the objective functions of policymakers differ based on the point estimates and the posterior distribution. That is, for any tax policy (μ, χ),

$$\underbrace{W(\mu, \chi, \mathbb{E}_{\Theta}[\Theta])}_{\text{based on point estimates}} \neq \underbrace{\mathbb{E}_{\Theta}[W(\mu, \chi, \Theta)]}_{\text{based on posterior distribution}} .$$

The left-hand side is what policy decisions are based on for policymakers treating the point estimates (i.e., posterior means) of parameters as their true values, whereas the right-hand side is for policymakers accounting for the parameter uncertainty correctly, i.e., the expected welfare gain with respect to the posterior distribution. In general, the two sides are not equal, and the difference depends on the convexity of the welfare gain function $W(\cdot)$ with respect to the uncertain parameters (similar to Jensen's inequality). Roughly speaking, the more convex is the welfare gain function, the larger is the expected welfare gain (the right-hand side) compared to the welfare gain based on the point estimates (the left-hand side).

However, such difference due to the convexity of the welfare gain function does not necessarily lead to different optimal policies. For example, if the expected welfare gain is always larger than the welfare gain based on the point estimates, but by a constant for any tax policy, the optimal policy would remain the same with or without the parameter uncertainty. Hence, for parameter uncertainty to change the optimal policy, the convexity of the welfare gain function must vary with the choice of tax policy.

Second, why does the uncertainty about preference and wage parameters lead to *more, not less,* progressive income tax? The short answer is that based on our quantitative model, the overall convexity of the welfare gain function with respect to the uncertain parameters *rises* with tax progressivity, which becomes an additional force that favors a more progressive income tax. The left panel of Figure 5 shows how the welfare gain based on the point estimates (the red dashed line) changes with tax progressivity. The curve is hump-shaped and reflects the standard trade-off faced by policymakers between the benefit of public insurance through progressive taxation and the efficiency cost of labor distortions. In the presence of parameter uncertainty, it becomes an additional consideration of policymakers, and its effect is captured by the difference between the expected welfare gain with respect to the posterior distribution (the blue solid line) and the welfare gain based on the point estimates. The right panel of Figure 5 shows that this difference is increasing in tax progressivity (i.e., the welfare gain function becomes more convex as tax progressivity rises), and as a result, it shifts the optimal tax progressivity to the right.

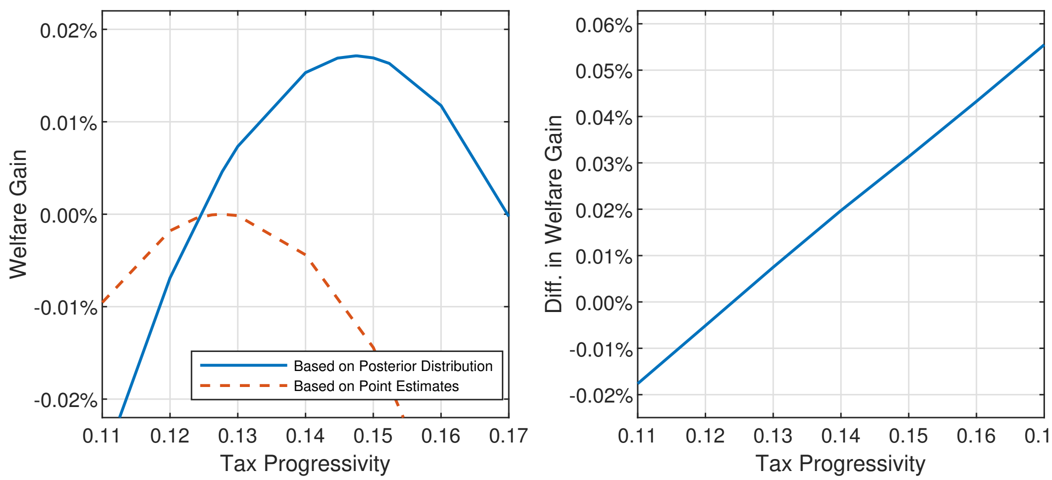

Figure 5: Tax Progressivity and Welfare

Notes: The left panel shows how the welfare gain relative to the status quo depends on the tax progressivity, based on the posterior distribution (blue solid line) and the point estimates (red dashed line). The right panel plots the difference in welfare gain between the two curves in the left panel. Welfare gain is reported as a percentage of lifetime consumption under the status quo tax policy and the point estimates of parameters.

To illustrate how tax progressivity affects the shape of the welfare gain function, Figure 6 plots the welfare gain function $W(\cdot)$ against three representative parameters under the status quo tax progressivity $\mu = 0.128$ (the red dashed line) and the optimal progressivity based on the posterior distribution $\mu = 0.147$ (the blue solid line). By construction, welfare gain is always zero under the status quo policy, and thus the welfare gain function is completely flat. However, as tax progressivity increases, the welfare gain function becomes convex in male wage persistence ρ_1, almost linear in male labor elasticity η_1, and concave in household risk aversion σ. Therefore, uncertainty about each of the three parameters *alone* has positive, zero, or negative effect on optimal tax progressivity. Conclusions are more obscure when we factor in the interactions and correlations between uncertain parameters. For example, although welfare gain appears to be almost linear in male labor elasticity in the center panel, this may no longer be true when other parameters are set at different values, and curvature may still exist along directions of correlations with other parameters.[26] The takeaway is that there is no universal answer about the effect of parameter uncertainty on tax progressivity. The answer depends on both the theoretical model specification and the empirical posterior distribution, and careful case-by-case quantitative examinations are more suitable for addressing such questions.

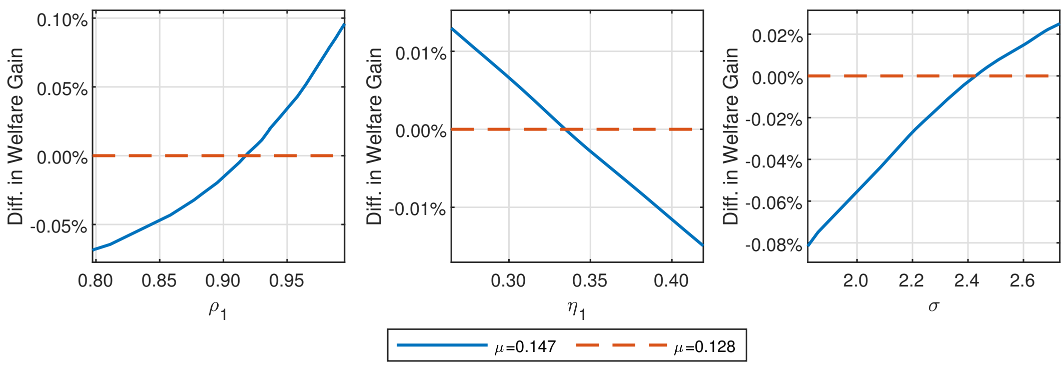

Figure 6: Welfare Gain and Uncertain Parameters

Notes: This figure plots the welfare gain function against the male wage persistence ρ_1 (left panel), male labor elasticitiy η_1 (center panel), and household risk aversion σ (right panel), when tax progressivity is increased from the status quo (red dashed line) to the optimal one based on the posterior distribution in Table 4 (blue solid line). In each plot, the other parameters are set at their point estimates, and the welfare gain at the point estimates is normalized to zero.

5.1.4 The Posterior Distribution Matters

One advantage of our structural Bayesian approach of dealing with parameter uncertainty over the standard sensitivity analysis in the literature is that our posterior distribution estimated from the data contains information about the correlations between uncertain parameters and the shape of their distribution. Through counter-

[26] A simple example is $f(x, y) = xy$. Notice that the function is linear in both x and y individually, but if x and y are perfectly positively correlated, the function is similar to x^2, which is convex. On the other hand, if x and y are perfectly negatively correlated, the function is similar to $-x^2$, which is concave. Again, we see the importance of correlations between parameters for analyses with parameter uncertainty.

factual experiments, we demonstrate in Table 5 that this information is crucial for quantitative evaluations of the effects of parameter uncertainty.

From the posterior distribution, we observe strong correlations (> 0.8 in absolute value) between parameters governing the levels of labor disutility (ψ_1, ψ_2, f) and other preference parameters. In the first counterfactual experiment, we artificially reverse the signs of these correlations while keeping the posterior means and variances of the marginal distributions untouched.[27] This change in correlation pattern leads to notable differences in uncertainty's effects on optimal tax policy: the rise in optimal tax progressivity due to parameter uncertainty is reduced by 25%, from 2 percentage points to 1.5 percentage points; and the reduction in optimal tax level is 33% smaller.

Table 5: Importance of the Posterior Distribution

Posterior ($\Pi(\Theta)$)	ΔProgressivity (pp)	ΔLevel (pp)
True Posterior	2.0	-0.6
Counterfactual		
Reversed Correlation	1.5	-0.4
Uniform Distribution	3.0	-0.8

Notes: The changes in optimal tax progressivity and level are reported in percentage point (pp). "Reversed Correlation" results are based on a posterior distribution with sign-reversed correlations between (ψ_1, ψ_2, f) and other parameters. "Uniform Distribution" results are based on a uniform posterior distribution with the same posterior means and variances.

In the second counterfactual experiment, we modify the shape of the marginal distributions of parameters to uniform distributions with the same means and variances, while maintaining the general correlation pattern between parameters.[28] This change of the posterior distribution again leads to significantly different evaluations about the uncertainty's effects on optimal tax policy: the optimal tax progressivity

[27] In particular, for ψ_1, ψ_2, and f, we replace each parameter value x with $2\mathbb{E}(x) - x$, where $\mathbb{E}(x)$ is the corresponding posterior mean.

[28] In particular, we replace each parameter value x with $\sqrt{3\text{Var}(x)}(2\text{cdf}(x) - 1) + \mathbb{E}(x)$, where $\mathbb{E}(x)$ and $\text{Var}(x)$ are the corresponding posterior mean and variance, and $\text{cdf}(x)$ is the cumulative distribution function of x. Note that the transformation is strictly increasing, and hence the correlation pattern between parameters is largely preserved, albeit not perfectly.

rises by 50% more than the benchmark result, and the optimal tax level declines by 33% more.

Since information about correlations between uncertain parameters and the shape of their distribution is often difficult to obtain through non-Bayesian methods, and it has been proven to be important through our counterfactual experiments, we propose that the Bayesian approach shall become the norm for quantitative studies in which parameter uncertainty is of primary concern.

5.1.5 Which Uncertainty is More Important?

To understand the importance of uncertainty about each aspect of the economy to the optimal design of tax policy, we conduct decomposition exercises based on our quantitative model. To keep the decomposition results informative and transparent, we categorize the uncertain parameters in two ways: i) those related to household preferences and those governing the idiosyncratic wage process; ii) male-related, female-related, and gender-neutral.[29]

Table 6: Decomposition: Uncertainty and Tax Progressivity

	ΔProgressivity (1)	Uncertainty (2)	Sensitivity (1)/(2)
Overall	100%	100%	1
A. Preferences vs. Wage Process			
Preference	-12%	30%	-0.40
Wage	112%	70%	1.60
B. Male vs. Female			
Male	90%	62%	1.45
Female	19%	11%	1.73
Neutral	-9%	27%	-0.33

Notes: "ΔProgressivity" denotes the change in (ex ante) optimal tax progressivity due to parameter uncertainty. "Uncertainty" is measured by the standard deviation of ex post optimal tax progressivity induced by parameter uncertainty. The results are normalized such that the total contribution equals 100%.

[29]For example, the male (female) labor supply elasticity η_1 (η_2) and wage persistence ρ_1 (ρ_2) are in the male (female) group, whereas the time discount rate δ and the correlation between male and female wage shocks $corr_{v_1,v_2}$ are considered gender-neutral.

The first column ("ΔProgressivity") of Table 6 reports the contribution of uncertainty about each parameter group to the change in optimal tax progressivity. We find that the increase in optimal tax progressivity due to parameter uncertainty in Table 4 is primarily driven by uncertainty about the idiosyncratic wage process (112%), and the effect of uncertainty about household preferences is much smaller and in the opposite direction (-12%). On the other hand, uncertainty about male-related parameters explains 90% of the change in optimal tax progressivity, and the contributions of uncertainties about female-related and gender-neutral parameters are 19% and -9%, respectively.

However, the results in the first column may not tell the full story since a larger effect of uncertainty about one group of parameters on optimal policy may be due to either i) there is a lot of uncertainty about these parameters, or ii) the optimal policy is more sensitive to uncertainty about these parameters. To separate these two channels, we need to first introduce a universal measure of uncertainty that is comparable across parameters with distinct economic interpretations. Given our focus on optimal tax progressivity, we choose to measure uncertainty by the standard deviation of *ex post* optimal tax progressivity induced by uncertainty about each group of parameters. More specifically, the ex post optimal tax policy is the optimal policy after parameter uncertainty is resolved, which is a function of parameter Θ and given by

$$(\widetilde{\mu}^*(\Theta), \widetilde{\chi}^*(\Theta)) = \arg\max_{(\mu,\chi)} W(\mu, \chi, \Theta).$$

Intuitively, if uncertainty about a parameter does not lead to significant variations in the ex post optimal policy, then no matter how dispersed is the posterior distribution of this parameter, there is not much uncertainty for policymakers about which policy to choose.

The second column ("Uncertainty") of Table 6 reports the decomposition of uncertainty based on this measure. While the effect of uncertainty about wage process on optimal tax progressivity is about nine times the effect of uncertainty about household preferences (112% vs. -12%) , uncertainty about wage process is only about twice as much as uncertainty about household preferences (70% vs. 30%). This suggests that optimal tax progressivity is more sensitive to uncertainty about wage process than household preferences. More specifically, the third column ("Sensitivity") simply computes the ratio of the first column to the second column, and we can see that the optimal tax progressivity is four times more sensitive to uncertainty about wage process (1.60 vs. –0.40). Therefore, uncertainty about wage process explains most of the change in optimal tax progresssivity not only because there is more uncertainty, but also because the optimal policy is more sensitive to it. On the other hand, while uncertainty about male-related parameters explains a majority of the change in optimal tax progressivity, the optimal policy is most sensitive to uncertainty about female-related parameters, followed by male-related and gender-neutral parameters.

5.2 Welfare Cost of Uncertainty

When policymakers are uncertain about the true values of key economic parameters, the ex ante optimal policy based on the posterior distribution in Section 5.1 is in general not optimal ex post. This leads to welfare losses compared to the first-best scenario in which policymakers have perfect information and can set the tax policy contingent on the values of parameters. We refer to such losses as the welfare cost of uncertainty since they can be eliminated by a signal that reveals the true values of parameters to the policymakers. It is useful to get a sense of how large this welfare cost is since there are potential ways of improving our knowledge about these parameters, yet at various degrees of costs. If the welfare cost of parameter uncertainty turns out to be smaller than the cost of eliminating such uncertainty, it would be rational for the policymakers to remain ignorant. If the opposite is true, more resources should be devoted to activities that can help us achieve better estimates of these parameters.

To quantify the welfare cost of uncertainty, let us first define the first-best tax policy which is a parameter-contingent policy plan that implements the ex post optimal tax policy state-by-state:[30]

$$(\widetilde{\mu}^*(\Theta), \widetilde{\chi}^*(\Theta)) = \arg\max_{(\mu,\chi)} W(\mu, \chi, \Theta),$$

[30] Here "state" refers to the state of nature that determines the values of parameters.

where $W(\mu, \chi, \Theta)$ is the same as in Section 5.1, and it represents the welfare gain from a tax reform relative to the status quo. The first-best tax policy maximizes the welfare for each parameter state, and hence it is the best the policymakers can do with perfect information about parameters. The welfare loss from adopting a parameter-invariant tax policy (μ, χ) relative to the first-best is then

$$\widehat{W}(\mu, \chi, \Theta) \equiv W(\widetilde{\mu}^*(\Theta), \widetilde{\chi}^*(\Theta), \Theta) - W(\mu, \chi, \Theta). \tag{7}$$

Notice that the welfare loss $\widehat{W}(\mu, \chi, \Theta)$ depends on the true parameter state Θ, and it is always nonnegative since no policy can induce higher welfare than the first-best policy.

The implementation of the first-best policy requires exact knowledge about the parameters, and hence it is infeasible when policymakers face parameter uncertainty. The best the policymakers can do in this case is to choose a parameter-invariant tax policy to minimize the welfare loss relative to the first-best. This gives rise to the following optimal tax policy problem:

$$(\mu^*, \chi^*) = \arg\min_{(\mu,\chi)} \Gamma^{-1}\left(\int_{\Theta} \Gamma\left(\widehat{W}(\mu, \chi, \Theta)\right) d\Pi(\Theta)\right), \tag{8}$$

where $\Gamma(\cdot)$ is again a strictly increasing function reflecting the risk preferences of policymakers, and $\Pi(\Theta)$ is the posterior distribution of parameters. For now, we maintain our previous assumption that policymakers are risk-neutral, i.e., $\Gamma(\cdot)$ is linear. And under this assumption, it is easy to verify that this problem of minimizing the welfare loss relative to the first-best is equivalent to the problem of maximizing the welfare gain relative to the status quo in Section 5.1. As a result, the ex ante optimal tax policy based on the posterior distribution is the same as in Table 4, i.e., $\mu^* = 0.147$, $\chi^* = 0.120$.

The left panel of Figure 7 plots the distribution of the ex post optimal tax progressivity implied by the posterior distribution of parameters. Depending on the parameter state, the ex post optimal tax progressivity varies considerably. However, since policymakers are uncertain about the true values of parameters, the ex ante optimal tax progressivity is fixed at $\mu^* = 0.147$, which leads to ex post welfare losses relative to the first-best outcome. The right panel of Figure 7 shows the distribution of this welfare loss, which is right-skewed with a lower bound of zero. The expected welfare loss is equivalent to 0.20% of lifetime consumption, which is about 29 billion dollars per year based on the 2019 U.S. consumption data.

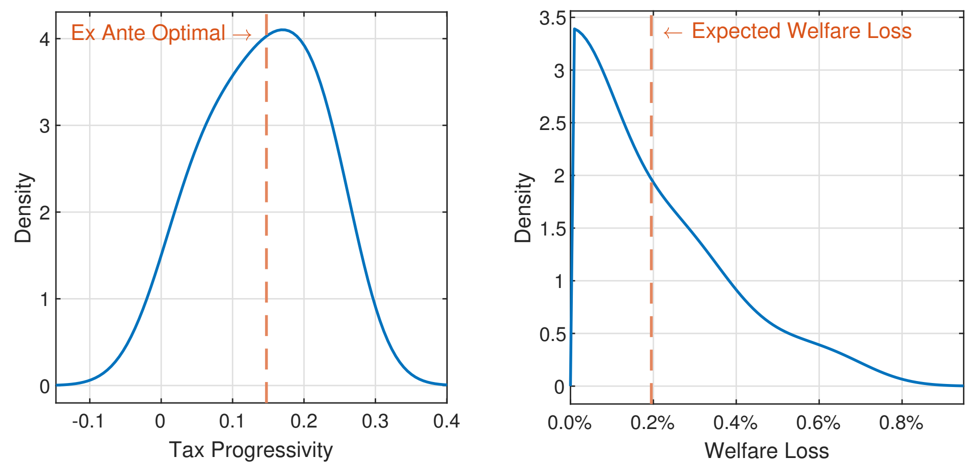

Figure 7: Distributions of the Ex Post Optimal Policy and Welfare Loss from Uncertainty

Notes: This figure shows the distributions of the ex post optimal tax progressivity (left panel) and the welfare loss relative to the first-best under the ex ante optimal tax policy (right panel). Welfare loss is reported as a percentage of lifetime consumption under the status quo tax policy and the point estimates of parameters.

It is worth stressing that the welfare cost of uncertainty we report only captures its distortionary effects on income tax policy. Since there are other government policy decisions that rely on information about household preferences and wage process, the total welfare cost of uncertainty through distortions to all policy decisions is likely much larger. Our finding thus indicates substantial potential welfare gains from activities that improve our knowledge about these aspects of the economy.

5.3 Risk Preferences of Policymakers

In the previous analyses, we have assumed that policymakers are risk-neutral with respect to the variations in welfare induced by parameter uncertainty. In this section, we investigate how our benchmark result may be affected by different risk preferences of policymakers.

For this purpose, it is important to differentiate two types of policymakers: i) "career politicians" who want to maximize the welfare gain relative to the status quo; and ii) "social planners" who want to minimize the welfare loss relative to the first-best. The optimal tax policy problems of these two types of policymakers are presented in (6) and (8), respectively. As mentioned in Section 5.2, these two types of policymakers are equivalent for optimal tax policy when they are risk-neutral, but it is no longer the case when policymakers are risk-averse. For example, maintaining the status quo policy is a risk-free option for "career politicians" since welfare gain is always zero. However, for "social planners", the status quo policy is risky in the sense that the welfare loss of no reform compared to the first-best is uncertain. Put another way, by doing nothing, "social planners" could be giving up very little or a lot in potential welfare gain depending on the true state of parameters. In the following, we consider both types of policymakers in our analysis.

Formally, we adjust the risk-preferences of policymakers by modifying the "utility function" of policymakers, i.e., $\Gamma(\cdot)$, in the optimal tax policy problem (6) and (8). In particular, we choose the CARA functional form for $\Gamma(\cdot)$,[31] and for "career politicians" in (6):

$$\Gamma(X) = \begin{cases} \frac{1-e^{-\alpha X}}{\alpha} & \text{if } \alpha \neq 0, \\ X & \text{if } \alpha = 0, \end{cases} \tag{9}$$

where α is the parameter that controls the risk aversion of policymakers. The larger is α, the more the policymakers dislike risk. When $\alpha = 0$, policymakers are risk-neutral, and we are back to the benchmark case. For "social planners' in (8), we only need to replace α with $-\alpha$ in (9).

Table 7 reports the optimal tax progressivity based on the posterior distribution under various degrees of policymakers' risk aversion and for both types of policymakers. We consider three degrees of risk aversion corresponding to $\alpha = 0$, 10, and 50. For "career politicians" (maximizing the welfare gain), the risk of tax reform is higher when the new policy is further away from the status quo policy, and hence risk aversion of policymakers pushes the optimal tax policy towards the status quo policy, which means lower progressivity in our model. For "social planners" (minimizing the welfare loss), the pattern is different. Since the first-best tax

[31] We choose the CARA, not CRRA, functional form because welfare gain, i.e., X in (9), could be negative.

policy depends on the true parameter values, a risk-free policy in this case must be parameter-contingent, which is infeasible in the presence of parameter uncertainty. Without a clear risk-free option, it is less straightforward how a higher risk aversion of policymakers would alter the optimal tax policy. Our quantitative results suggest that the optimal tax progressivity is actually quite robust with respect to the degree of policymakers' risk aversion.

Table 7: Optimal Tax Progressivity and Risk Preferences of Policymakers

	Career Politicians (Max Welfare Gain)	Social Planners (Min Welfare Loss)
Risk Neutral	0.147	0.147
Risk Aversion 10	0.143	0.147
Risk Aversion 50	0.135	0.147
Tail 10	0.128	0.145

Notes: "Risk Neutral", "Risk Aversion 10", and "Risk Aversion 50" correspond to the cases with $\alpha = 0$, 10, and 50. "Tail 10" corresponds to the case when the average of the worst 10% of possible outcomes is used as the objective.

Figure 8 shows how welfare measures change with tax progressivity for both types of policymakers with different degrees of risk aversion. In addition to reconfirming our findings in Table 7 about the optimal policy (i.e., the peaks and troughs of curves), one interesting observation is that for "career politicians", the welfare gain from any tax reform that increases tax progressivity relative to the status quo is lowered by the policymakers' risk aversion, whereas for "social planners", the welfare gain of such reform (i.e., reduction in welfare loss) increases with policymakers' risk aversion. In other words, more risk-averse "career politicians" have weaker incentives to conduct progressive tax reforms, whereas the opposite is true for "social planners".

We also consider another form of policymakers' risk preferences focusing on the worst 10% of possible outcomes. That is, for "career politicians", we assume that the policymakers want to maximize the average of the bottom 10% of the welfare distribution induced by parameter uncertainty. For "social planners", it would be minimizing the top 10% of the welfare loss distribution. This specification is, in spirit, similar to the max-min criterion from the robust control literature. The cor-

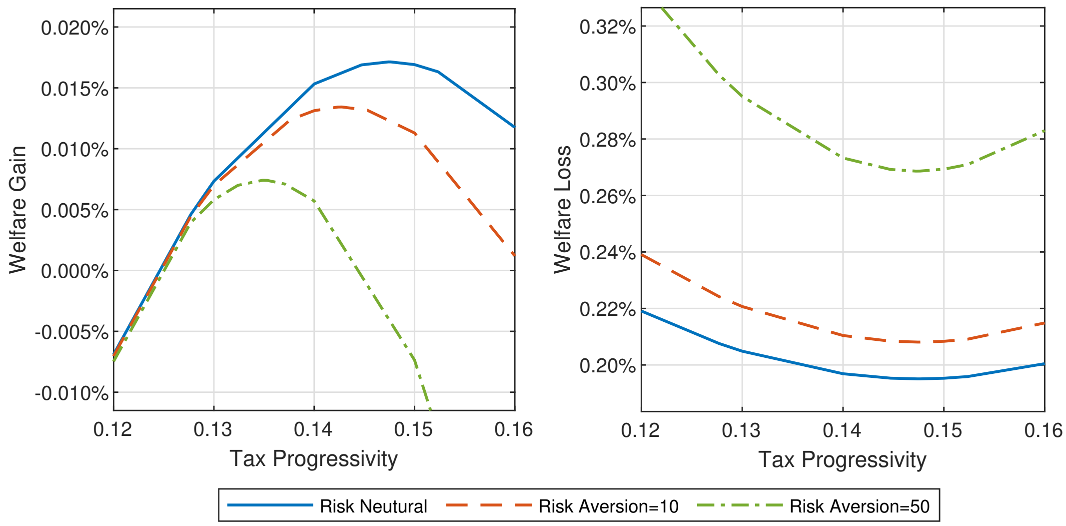

Figure 8: Welfare, Tax Progressivity, and Risk Preferences of Policymakers

Notes: This figure shows how welfare measures change with tax progressivity, for "career politicians" (left panel) and "social planners" (right panel) of various degrees of risk aversion. "Risk Neutral", "Risk Aversion 10", and "Risk Aversion 50" correspond to the cases with $\alpha = 0$, 10, and 50.

responding results are reported in Table 7 under the label "Tail 10". The tax policy implications of this type of risk preferences are quite similar to those of the standard risk aversion controlled by α. Compared to the risk-neutral case, for "career politicians", it reduces the optimal tax progressivity significantly (0.128 vs. 0.147); for "social planners", however, the optimal tax progressivity stays almost the same (0.145 vs. 0.147).

6 Conclusions

In this paper, we have examined the income tax policy implications of information frictions in the form of policymakers' uncertainty about parameters governing household preferences and wage process. We show in theory that the effect of uncertainty on optimal tax progressivity depends on how the curvature of the welfare measure with respect to the uncertain parameters is affected by tax progressivity. To quantify the effect, we employ a limited-information Bayesian approach to measure uncertainty in the form of posterior distribution from the U.S. data, and conduct thought experiments of tax reform in an incomplete-market life-cycle model of heterogeneous households.

Our quantitative analysis suggests that the existence of such uncertainty leads to more progressive optimal income tax, which would raise the gap in marginal tax rate between high- and low-income households by 5 percentage points. The increase in optimal tax progressivity is primary driven by uncertainty about wage process, whereas uncertainty about household preferences has small and opposite effect. The posterior distribution plays an important role since it contains key information about the correlations between uncertain parameters and the shape of their distribution.

Our approach of quantifying uncertainty from the data and analyzing its implications on policy design and welfare evaluation is universal and not unique to uncertainty about household preferences and wage process or the income tax policy. Therefore, a natural next step for future research would be to apply the general approach of this paper to uncertainty about other aspects of the economy and related government policies. For computational reasons, we have assumed that policymakers are passive learners in the sense that they take the data they observe as given. It would be interesting to explore what would happen if active learning is introduced such that policymakers can take actions that directly or indirectly affect the information they receive, although the model may have to be simplified substantially along other dimensions to make such extension feasible.

References

Bakış, Ozan, Barış Kaymak, and Markus Poschke (2015). "Transitional Dynamics and the Optimal Progressivity of Income Redistribution". *Review of Economic Dynamics* 18.3, pp. 679–693.

Blundell, Richard, Luigi Pistaferri, and Itay Saporta-Eksten (2016). "Consumption Inequality and Family Labor Supply". *American Economic Review* 106.2, pp. 387–435.

Brainard, William C. (1967). "Uncertainty and the Effectiveness of Policy". *The American Economic Review* 57.2, pp. 411–425.

Bénabou, Roland (2002). "Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?" *Econometrica* 70.2, pp. 481–517.

Chang, Bo Hyun, Yongsung Chang, and Sun-Bin Kim (2018). “Pareto weights in practice: A quantitative analysis across 32 OECD countries”. *Review of Economic Dynamics* 28, pp. 181–204.

Christiano, Lawrence J., Mathias Trabandt, and Karl Walentin (2010). “Chapter 7 - DSGE Models for Monetary Policy Analysis”. Ed. by Benjamin M. Friedman and Michael Woodford. Vol. 3. Handbook of Monetary Economics. Elsevier, pp. 285–367.

Conesa, Juan Carlos and Dirk Krueger (2006). “On the Optimal Progressivity of the Income Tax Code”. *Journal of Monetary Economics* 53.7, pp. 1425–1450.

Conesa, Juan Carlos, Sagiri Kitao, and Dirk Krueger (2009). “Taxing Capital? Not a Bad Idea after All!” *American Economic Review* 99.1, pp. 25–48.

Edge, Rochelle M., Thomas Laubach, and John C. Williams (2010). “Welfare-Maximizing Monetary Policy under Parameter Uncertainty”. *Journal of Applied Econometrics* 25.1, pp. 129–143.

Fernández-Villaverde, J., J.F. Rubio-Ramírez, and F. Schorfheide (2016). “Chapter 9 - Solution and Estimation Methods for DSGE Models”. Ed. by John B. Taylor and Harald Uhlig. Vol. 2. Handbook of Macroeconomics. Elsevier, pp. 527–724.

Hansen, Lars Peter and Thomas J. Sargent (2008). *Robustness*. Princeton University Press.

Heathcote, Jonathan and Hitoshi Tsujiyama (2021). “Optimal Income Taxation: Mirrlees Meets Ramsey”. *Journal of Political Economy* 129.11, pp. 3141–3184.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante (2017). “Optimal Tax Progressivity: An Analytical Framework”. *The Quarterly Journal of Economics* 132.4, pp. 1693–1754.

Heckman, James, Lance Lochner, and Ricardo Cossa (2002). *Learning-By-Doing Vs. On-the-Job Training: Using Variation Induced by the EITC to Distinguish Between Models of Skill Formation*. Working Paper 9083. National Bureau of Economic Research.

Kim, Jae-Young (2002). “Limited Information Likelihood and Bayesian Analysis”. *Journal of Econometrics* 107.1, pp. 175–193.

Kimball, Miles S. (1990). “Precautionary Saving in the Small and in the Large”. *Econometrica* 58.1, pp. 53–73.

Krueger, Dirk and Alexander Ludwig (2016). “On the Optimal Provision of Social Insurance: Progressive Taxation versus Education Subsidies in General Equilibrium”. *Journal of Monetary Economics* 77, pp. 72–98.

Levin, Andrew T., Alexei Onatski, John C. Williams, and Noah Williams (2005). “Monetary Policy under Uncertainty in Micro-Founded Macroeconometric Models”. *NBER Macroeconomics Annual* 20, pp. 229–287.

Lockwood, Benjamin B., Afras Sial, and Matthew Weinzierl (2020). “Designing, Not Checking, for Policy Robustness: An Example with Optimal Taxation”. *Tax Policy and the Economy, Volume 35*. University of Chicago Press, pp. 1–54.

Online Appendix

When in Doubt, Tax More Progressively? Uncertainty and Progressive Income Taxation

Minsu Chang and Chunzan Wu

A Supplementary Theoretical Results

A.1 Welfare Gain from Tax Reform

In this section, we derive the closed-form formula for the welfare gain from tax reform $W(\mu, \Theta)$ of Section 2.2 in greater detail. Given the income tax function, maximizing household utility subject to the budget constraint yields the optimal household labor supply

$$H(z;\mu,\chi,\Theta) = (1-\mu)^{\frac{1}{1+\eta^{-1}}}.$$

Note that household labor supply does not depend on productivity z or tax level χ, and households treat public consumption G as exogenous when making labor supply decisions. Private consumption of productivity z household is then simply the after-tax income

$$C(z;\mu,\chi,\Theta) = (1-\chi)[zH(z;\mu,\chi,\Theta)]^{1-\mu}.$$

Based on the government budget constraint, we can express government spending on public consumption as

$$\begin{aligned} G(\mu,\chi,\Theta) =& \int_z zH(z;\mu,\chi,\Theta)dF_z(z) - (1-\chi)\int_z [zH(z;\mu,\chi,\Theta)]^{1-\mu}\, dF_z(z) \\ =&(1-\mu)^{\frac{1}{1+\eta^{-1}}} - (1-\chi)(1-\mu)^{\frac{1-\mu}{1+\eta^{-1}}} \underbrace{\int_z z^{1-\mu} dF_z(z)}_{=\mathbb{E}[z^{1-\mu}]}. \end{aligned}$$

Suppose that the policymakers are utilitarian. The social welfare under parameter $\Theta = (\eta, \sigma_z)$ is simply

$$\mathrm{SWF}(\mu, \chi, \Theta) = \int_z \left[\ln C(z; \mu, \chi, \Theta) - \frac{H(z; \mu, \chi, \Theta)^{1+\eta^{-1}}}{1+\eta^{-1}} + \gamma \ln G(\mu, \chi, \Theta) \right] dF_z(z).$$

Plugging in the consumption and labor supply policy functions and integrating over the distribution of household productivity, we have

$$\mathrm{SWF}(\mu, \chi, \Theta) = \ln(1-\chi) - \frac{(1-\mu)}{2}\sigma_z^2 + \frac{1-\mu}{1+\eta^{-1}} \ln(1-\mu) - \frac{1-\mu}{1+\eta^{-1}} + \gamma \ln G(\mu, \chi, \Theta).$$

For tractability, we assume that tax level χ is set contingent on tax progressivity μ and parameter Θ to maximize the ex post social welfare. The ex post optimality condition for χ is then

$$\frac{1}{1-\chi} = \frac{\gamma}{G(\mu, \chi, \Theta)} \frac{\partial G(\mu, \chi, \Theta)}{\partial \chi},$$

which implies that the ex post optimal tax level is

$$\chi^*(\mu, \Theta) = 1 - \frac{(1-\mu)^{\frac{\mu}{1+\eta^{-1}}}}{(1+\gamma)\mathbb{E}[z^{1-\mu}]} = 1 - \frac{(1-\mu)^{\frac{\mu}{1+\eta^{-1}}}}{(1+\gamma)} e^{\frac{\mu(1-\mu)\sigma_z^2}{2}}.$$

Note that when $\mu = 0$, the ex post optimal tax level is always $\chi = \frac{\gamma}{1+\gamma}$, independent of parameter Θ. The ex post optimal amount of public consumption is then

$$G(\mu, \chi^*(\mu, \Theta), \Theta) = \frac{\gamma}{1+\gamma}(1-\mu)^{\frac{1}{1+\eta^{-1}}}.$$

Let (μ_{sq}, χ_{sq}) denote the status quo tax policy. If policymakers adopt a new tax policy with progressivity μ and the corresponding ex post optimal tax level $\chi^*(\mu, \Theta)$, the social welfare gain from such tax reform is then

$$\Delta(\mu, \Theta) \equiv \mathrm{SWF}(\mu, \chi^*(\mu, \Theta), \Theta) - \mathrm{SWF}(\mu_{sq}, \chi_{sq}, \Theta).$$

When $\mu_{sq} = 0$ and $\chi_{sq} = \frac{\gamma}{1+\gamma}$, we have

$$\Delta(\mu, \Theta) = \frac{1}{2}\sigma_z^2 \mu(2-\mu) + \frac{1}{1+\eta^{-1}} \left[(1+\gamma)\ln(1-\mu) + \mu \right].$$

To convert the welfare gain $\Delta(\mu, \Theta)$ into consumption equivalent variation $W(\mu, \Theta)$, we first need to find the percentage change in consumption x that would generate the same welfare change as the tax reform, i.e.,

$$\Delta(\mu, \Theta) = \int_z [\ln((1+x)C(z; \mu_{sq}, \chi_{sq}, \Theta)) - \ln C(z; \mu_{sq}, \chi_{sq}, \Theta)] \, dF_z(z) = \ln(1+x),$$

which implies

$$x = e^{\Delta(\mu,\Theta)} - 1.$$

Hence the total amount of consumption transfers is

$$W(\mu, \Theta) = x \int_z C(z; \mu_{sq}, \chi_{sq}, \Theta) dF_z(z) = \left(e^{\Delta(\mu,\Theta)} - 1\right) \frac{1}{1+\gamma}.$$

A.2 Signs of Derivatives

For the static model in Section 2.2, we have the following derivatives of $\Delta(\mu, \Theta)$ with respect to the Frisch elasticity η:

$$\begin{aligned}
\Delta_\eta &= \eta^{-2}(1+\eta^{-1})^{-2} \left[(1+\gamma)\ln(1-\mu) + \mu\right]; \\
\Delta_{\eta\eta} &= -2\eta^{-3}(1+\eta^{-1})^{-3} \left[(1+\gamma)\ln(1-\mu) + \mu\right]; \\
\Delta_{\eta\mu} &= \eta^{-2}(1+\eta^{-1})^{-2} \left(1 - \frac{1+\gamma}{1-\mu}\right); \\
\Delta_{\eta\eta\mu} &= -2\eta^{-3}(1+\eta^{-1})^{-3} \left(1 - \frac{1+\gamma}{1-\mu}\right).
\end{aligned}$$

Since $W(\mu, \Theta) = \left(e^{\Delta(\mu,\Theta)} - 1\right)/(1+\gamma)$, we have

$$W_{\eta\eta\mu} = \frac{1}{1+\gamma} \left\{ e^{\Delta} \left[\Delta_{\eta\eta\mu} + 2\Delta_\eta \Delta_{\eta\mu}\right] + e^{\Delta} \Delta_\mu \left[\Delta_{\eta\eta} + (\Delta_\eta)^2\right] \right\}.$$

Since $\bar{\mu}$ maximizes $W(\mu, \bar{\Theta})$ and thus $\Delta(\mu, \bar{\Theta})$ by definition, we have $\Delta_\mu(\bar{\mu}, \bar{\Theta}) = 0$. It follows that when $\bar{\mu} \in (0, 1)$ and $\gamma > 0$,

$$W_{\eta\eta\mu}(\bar{\mu}, \bar{\Theta}) = \frac{1}{1+\gamma} e^{\Delta(\bar{\mu},\bar{\Theta})} [\underbrace{\Delta_{\eta\eta\mu}(\bar{\mu}, \bar{\Theta})}_{>0} + 2 \underbrace{\Delta_\eta(\bar{\mu}, \bar{\Theta})}_{<0} \underbrace{\Delta_{\eta\mu}(\bar{\mu}, \bar{\Theta})}_{<0}] > 0.$$

Similarly, we have the following derivatives of $\Delta(\mu, \Theta)$ with respect to the magnitude of idiosyncratic productivity risk σ_z:

$$\begin{aligned}
\Delta_{\sigma_z} &= \sigma_z \mu(2-\mu), \\
\Delta_{\sigma_z\sigma_z} &= \mu(2-\mu), \\
\Delta_{\sigma_z\mu} &= 2\sigma_z(1-\mu), \\
\Delta_{\sigma_z\sigma_z\mu} &= 2(1-\mu),
\end{aligned}$$

and

$$W_{\sigma_z\sigma_z\mu} = \frac{1}{1+\gamma}\left\{e^{\Delta}[\Delta_{\sigma_z\sigma_z\mu} + 2\Delta_{\sigma_z}\Delta_{\sigma_z\mu}] + e^{\Delta}\Delta_{\mu}[\Delta_{\sigma_z\sigma_z} + (\Delta_{\sigma_z})^2]\right\}.$$

Notice again that $\Delta_\mu(\bar{\mu}, \bar{\Theta}) = 0$, and hence when $\bar{\mu} \in (0, 1)$ and $\gamma > 0$,

$$W_{\sigma_z\sigma_z\mu}(\bar{\mu}, \bar{\Theta}) = \frac{1}{1+\gamma} e^{\Delta(\bar{\mu}, \bar{\Theta})}[\underbrace{\Delta_{\sigma_z\sigma_z\mu}(\bar{\mu}, \bar{\Theta})}_{>0} + 2\underbrace{\Delta_{\sigma_z}(\bar{\mu}, \bar{\Theta})}_{>0}\underbrace{\Delta_{\sigma_z\mu}(\bar{\mu}, \bar{\Theta})}_{>0}] > 0.$$

A.3 Exploration with Heathcote et al. (2017)'s Model

Heathcote et al. (2017) (hereafter HSV) develop a tractable model to study the optimal income tax problem with a closed-from expression for social welfare. In this section, we attempt to extend their model to include parameter uncertainty, and unfortunately, we find that it is generally difficult to obtain a closed-form expression for the expected welfare that is needed for analyses with parameter uncertainty. The main difficulty is that social welfare is highly nonlinear in the uncertain parameters, and integration of such nonlinear function for the expected welfare often does not allow a closed-from expression.

To make the point, it is enough to consider a stripped-down version of the HSV model. In particular, we shut down the endogenous skill investment and preference heterogeneity in their model, and keep their original notation for ease of comparison. The closed-form social welfare $\mathcal{W}(g,\tau)$ then becomes:[32]

$$\begin{aligned}\mathcal{W}(g,\tau) = & \log(1-g) + \chi \log g + (1+\chi)\frac{\log(1-\tau)}{(1+\hat{\sigma})(1-\tau)} - \frac{1}{1+\hat{\sigma}} \\ & - (1-\tau)^2 \frac{\upsilon_\omega}{2} \\ & + (1+\chi)\left[\frac{1}{\hat{\sigma}}\upsilon_\epsilon - \sigma\frac{1}{\hat{\sigma}^2}\frac{\upsilon_\epsilon}{2}\right].\end{aligned}$$

Here g is the fraction of output devoted to public consumption, and τ is the degree of tax progressivity. $\hat{\sigma} = \frac{\sigma+\tau}{1-\tau}$ where $\frac{1}{\hat{\sigma}}$ and $\frac{1}{\sigma}$ are the tax-modified and unmodified labor elasticities, respectively. υ_ω is the cross-sectional variance of the uninsurable wage shocks α, and υ_ϵ is the variance of the insurable wage shocks ϵ. χ is the taste parameter for the public good G. The sum of the first four terms is the same as the social welfare from a representative agent model. The fifth term is associated with uninsurable shocks, and the remaining terms are related to insurable shocks.

When there is parameter uncertainty, as explained in our main text, we need to first convert welfare gains from tax form into consumption equivalent variations such that they can be meaningfully compared across parameter states. With log utility in consumption, the welfare gain of adopting policy (g,τ) in comparison to the baseline policy (g_0,τ_0) is, in consumption change,

$$\left[\exp(\mathcal{W}(g,\tau) - \mathcal{W}(g_0,\tau_0)) - 1\right] C(g_0,\tau_0),$$

where $C(g_0,\tau_0)$ is the total consumption under policy (g_0,τ_0).

From Corollary 6 of HSV, the welfare-maximizing g is given by $g = \frac{\chi}{1+\chi}$. Suppose there is no uncertainty in χ, then g can always be set at the welfare-maximizing level without the exact knowledge about other parameters. Hence, with some abuse of notation, we denote the social welfare as a function of τ hereafter, $\mathcal{W}(\tau)$. Let the

[32] The formula is based on Proposition 4 of Heathcote et al. (2017). When we abstract from endogenous skill investment and from preference heterogeneity, equation (30) in their paper can be simplified. Term (b), (c), (d) are associated with endogenous skill investment, and (e) is with preference heterogeneity. Furthermore, we assume $\gamma = \beta$ so that term (f) is approximated by $-(1-\tau)^2\frac{\upsilon_\omega}{2}$.

status quo tax progressivity be $\tau_0 = 0$, and we have

$$\begin{aligned}
\exp(\mathcal{W}(\tau)) =&(1-g)\cdot g^{\chi}\cdot(1-\tau)^{\frac{(1+\chi)}{(1+\hat{\sigma})(1-\tau)}}\cdot\exp\left(-\frac{1}{1+\hat{\sigma}}\right)\\
&\times\exp\left(-(1-\tau)^2\frac{\upsilon_\omega}{2}\right)\cdot\exp\left((1+\chi)\left[\frac{1}{\hat{\sigma}}\upsilon_\epsilon-\sigma\frac{1}{\hat{\sigma}^2}\frac{\upsilon_\epsilon}{2}\right]\right)\\
\exp(\mathcal{W}(0)) =&(1-g)\cdot g^{\chi}\cdot\exp\left(-\frac{1}{1+\sigma}\right)\\
&\times\exp\left(-\frac{\upsilon_\omega}{2}\right)\cdot\exp\left((1+\chi)\frac{1}{\sigma}\frac{\upsilon_\epsilon}{2}\right)\\
\exp(\mathcal{W}(\tau)-\mathcal{W}(0)) =&(1-\tau)^{\frac{(1+\chi)}{(1+\sigma)}}\cdot\exp\left(\frac{\tau}{\sigma+1}\right)\\
&\times\exp\left((2-\tau)\tau\frac{\upsilon_\omega}{2}\right)\cdot\exp\left((1+\chi)\left[\frac{1}{\hat{\sigma}}-\frac{\sigma}{2\hat{\sigma}^2}-\frac{1}{2\sigma}\right]\upsilon_\epsilon\right)
\end{aligned}$$

The expected welfare gain is the integration of the welfare gain in consumption with respect to the posterior distribution of uncertain parameters (e.g., σ):

$$\mathbb{E}\left[\left(\exp(\mathcal{W}(\tau)-\mathcal{W}(0))-1\right)C(0)\right],$$

where $C(0) = \frac{1}{1+\chi}\exp\left(\frac{\upsilon_\epsilon}{2\sigma}\right)$, and we can already see that it is in general difficult to obtain a closed-form formula.[33]

B Supplementary Materials for Quantitative Analysis

B.1 Additional Quantitative Results

In this section, we provide supplementary results to the quantitative analysis in the main text. Figure 9 presents the estimated male and female log-wage trends over the life cycle. Table 8 reports the list of moment conditions employed in the limited-information Bayesian estimation and their values from the data and from the posterior distribution. Table 9 and 10 report the correlations among preference

[33]We can work out a case where a closed form of the expected welfare gain is available when υ_ω is the only uncertain parameter and its uncertainty can be characterized by a uniform distribution (available upon request). However, in general, getting a closed form for the integral of the welfare gain given uncertain parameters' distribution is difficult.

and wage parameters, respectively.

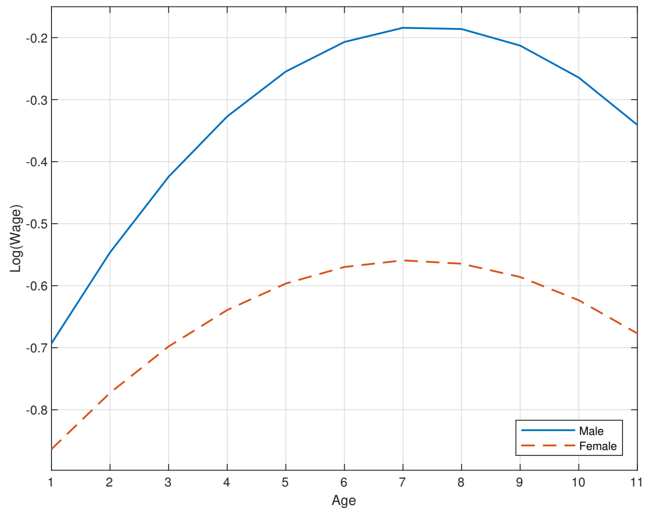

Figure 9: Male and Female Wage Trends

Notes: This figure plots the male (blue solid line) and female (red dashed line) log-wage trends over the life cycle estimated from the data.

Table 8: Moment Conditions

Moment	Data	Model
A. Wage Parameters		
$\mathbb{E}(\Delta w_{1,t}\Delta w_{1,t-1})$	-0.0264	-0.0272
$\mathbb{E}[(\Delta w_{1,t})^2]$	0.1332	0.1331
$\mathbb{E}(\Delta w_{2,t}\Delta w_{2,t-1})$	-0.0352	-0.0348
$\mathbb{E}[(\Delta w_{2,t})^2]$	0.1487	0.1480
$\mathbb{E}(\Delta w_{1,t}\Delta w_{2,t})$	0.0062	0.0061
B. Preference Parameters		
$\mathbb{E}[(\Delta c_t)^2]$	0.0711	0.0479
$\mathbb{E}(\Delta w_{1,t}\Delta y_{1,t})$	0.0988	0.1170
$\mathbb{E}(\Delta w_{2,t}\Delta y_{2,t})$	0.0777	0.1339
$\mathbb{E}[(\Delta y_{1,t})^2]$	0.1275	0.1269
$\mathbb{E}[(\Delta y_{2,t})^2]$	0.1962	0.1571
$\mathbb{E}(\Delta y_{1,t}\Delta y_{1,t-1})$	-0.0264	-0.0176
$\mathbb{E}(\Delta y_{2,t}\Delta y_{2,t-1})$	-0.0211	-0.0291
$\mathbb{E}(\Delta y_{1,t}\Delta c_t)$	0.0078	0.0234
$\mathbb{E}(\Delta y_{2,t}\Delta c_t)$	0.0022	0.0123
$\mathbb{E}(\Delta w_{1,t}\Delta c_t)$	0.0059	0.0233
$\mathbb{E}(\Delta w_{2,t}\Delta c_t)$	-0.0014	0.0142
$\mathbb{E}(\Delta y_{1,t}\Delta y_{2,t})$	-0.0001	-0.0170
$\mathbb{E}(\Delta w_{1,t}\Delta y_{2,t})$	-0.0008	-0.0039
$\mathbb{E}(\Delta w_{2,t}\Delta y_{1,t})$	0.0044	-0.0017
$\mathbb{E}[Y_{1,t}]$	1	0.9412
$\mathbb{E}[Y_{2,t}\|H_{2,t} > 0]$	0.4809	0.4677
$\mathbb{E}[\mathbf{I}(H_{2,t} = 0)]$	0.1539	0.1574
$\mathbb{E}[H_{1,t}]$	1	1.0019
$\mathbb{E}[H_{2,t}\|H_{2,t} > 0]$	0.7058	0.7262
$\mathbb{E}[A_t]$	0.8836	0.9021

Notes: "Model" results are the means of the posterior distributions of moments.

Table 9: Correlations Between Preference Parameters

	σ	η_1	η_2	ψ_1	ψ_2	f	δ
σ	1	0.21	-0.22	0.95	0.47	0.44	-0.46
η_1	0.21	1	-0.35	0.14	0.36	0.27	0.08
η_2	-0.22	-0.35	1	-0.20	-0.92	-0.87	0.01
ψ_1	0.95	0.14	-0.20	1	0.47	0.46	-0.48
ψ_2	0.47	0.36	-0.92	0.47	1	0.90	-0.16
f	0.44	0.27	-0.87	0.46	0.90	1	-0.16
δ	-0.46	0.08	0.01	-0.48	-0.16	-0.16	1

Table 10: Correlations Between Wage Parameters

	ρ_1	ρ_2	$\sigma^2_{v_1}$	$\sigma^2_{v_2}$	$corr_{v_1,v_2}$
ρ_1	1	0.05	-0.48	0.04	0.14
ρ_2	0.05	1	-0.07	0.15	-0.06
$\sigma^2_{v_1}$	-0.48	-0.07	1	0.02	-0.07
$\sigma^2_{v_2}$	0.04	0.15	0.02	1	-0.01
$corr_{v_1,v_2}$	0.14	-0.06	-0.07	-0.01	1

B.2 Measure Welfare Change in Consumption

We provide here the formulas for computing the welfare gain in consumption $W(\mu, \chi, \Theta)$ in the optimal tax policy problem of Section 5.1.1. Following the convention, we assume that the hypothetical consumption transfers are proportional to household consumption before the tax reform, while household labor supply and public consumption remain the same. Given the CRRA utility function in private consumption, the proportional change in consumption required to achieve the same social welfare change as the tax reform is given by

$$\text{CEV}(\mu, \chi, \Theta) \equiv \left(1 + \frac{\text{SWF}(\mu, \chi, \Theta) - \text{SWF}(\mu_{sq}, \chi_{sq}, \Theta)}{\text{VC}(\mu_{sq}, \chi_{sq}, \Theta)}\right)^{1/(1-\sigma)} - 1.$$

Here $\text{VC}(\mu_{sq}, \chi_{sq}, \Theta)$ is the weighted sum of expected lifetime utility from private consumption for a newborn cohort under the status quo tax policy (μ_{sq}, χ_{sq}). The weights are the same as the Pareto weights $\omega(\mathbf{s})$ in the social welfare function.

The present value of all consumption transfers required is then:

$$W(\mu, \chi, \Theta) \equiv \text{CEV}(\mu, \chi, \Theta) \times \left[\sum_{t=1}^{T} \frac{\text{Consumption}_t(\mu_{sq}, \chi_{sq}, \Theta)}{(1+r)^{t-1}}\right],$$

$$\text{Comsumption}_t(\mu_{sq}, \chi_{sq}, \Theta) = \int_{\mathbf{s}} C_t(\mathbf{s}; \mu_{sq}, \chi_{sq}, \Theta) d\Phi_t(\mathbf{s}; \mu_{sq}, \chi_{sq}, \Theta),\ t = 1, \ldots, T,$$

where $C_t(\mathbf{s}; \mu_{sq}, \chi_{sq}, \Theta)$ is household consumption, $\Phi_t(\mathbf{s}; \mu_{sq}, \chi_{sq}, \Theta)$ is the distribution of households, and $\text{Consumption}_t(\mu_{sq}, \chi_{sq}, \Theta)$ is total private consumption, all for age-t households under the status quo tax policy.

B.3 Computation

The household optimization problem is solved backwards using the endogenous grid method. With the extensive margin of female labor supply, for each iteration and each household state, the optimization problem is solved twice under two alternative scenarios: the current period female labor supply is strictly positive or zero. The final optimal policy is obtained by comparing the discounted utility achieved in these two scenarios.

The grid for asset has 100 grid points, and the distance between two adjacent grid points increases with the asset level such that the grid points are denser around the low asset levels where borrowing constraints are more likely to bind. The joint process of the two earners' idiosyncratic wage components is approximated by a discrete Markov process with age-dependent sets of states and transition matrices. The grid for the joint wage process has 11 points in each dimension, so there are in total 121 grid points at each age.

Since the computation burden of the optimal tax policy problem with parameter uncertainty is proportional to the number of posterior draws, we cannot afford using the empirical posterior distribution with 15,000 draws directly.[34] Hence to keep the computation burden manageable, we approximate the empirical posterior distribution with a sample of 30 i.i.d draws from it. We adjust the values of these draws such that the means and variances of the approximated distribution are exactly the same as those of the true posterior distribution. In particular, we replace the parameter value $\hat{\theta}$ of each draw by

$$\frac{\sigma_\theta}{\sigma_{\hat{\theta}}}(\hat{\theta} - \mu_{\hat{\theta}}) + \mu_\theta,$$

where μ_θ and σ_θ are the mean and standard deviation of parameter θ implied by the true posterior distribution; and $\mu_{\hat{\theta}}$ and $\sigma_{\hat{\theta}}$ are the sample mean and standard deviation of the unadjusted approximated distribution. Since the means and variances of the unadjusted distribution are already close to those of the true posterior distribu-

[34]Notice that previous quantitative studies of optimal tax policy in the literature are equivalent to solving the problem with only 1 posterior draw. So the increase in computation burden here is non-trivial.

tion, the adjustments required to make them equal are rather minor. For robustness check, we double the number of sample draws for the approximated distribution and redo our baseline exercise: the change in optimal tax progressivity based on the approximated posterior distribution is less than 1e-3. Therefore, we conclude that increasing the number of sample draws further would not affect the results significantly.