# Estimate the Belief Bias in Learning from Coworkers

Shaoshuang Yang*
University of Southern California

Preliminary Draft
This version:October 4, 2022
Click here for the latest version

**Abstract**

This paper studies the belief bias in the workplace. I build a structural model where workers can learn from coworkers. They choose where to work based on both wage and perceived learning opportunities. I estimate the model using German administrative data which contain information on workforce composition and workers' characteristics. I propose a methodology to separately estimate the perceived and the correct learning functions, building on the observation that learning is priced by a competitive market based on belief. The estimation results show that workers overestimate how much they can learn from coworkers by seven times. It implies that better knowledgeable workers are overpaid while the rest are underpaid, which increases the within-team inequality.

Keywords: Belief, knowledge diffusion, growth, income distribution, peer effects.

*shaoshuy@usc.edu.

# 1 Introduction

Do we learn as much as we expected? This paper structurally estimates the belief deviation from the correct probabilities in a learning-from-coworker setting. In reality, "you can learn from experienced colleagues" is often seen in a job advertisement, and people do value it. One may agree to take a low wage just for the learning opportunities, for example, in apprenticeship. The question is, is it worth it? Will the future wage be high enough to compensate for the sacrifices today?

This question is challenging because we can observe neither people's beliefs nor the sorting in the hiring procedure. I develop a novel way to disentangle the belief from reality. I generalize the model proposed by Jarosch et al. (2021) by allowing the workers have biased belief about how much they can learn from coworkers. In the model, workers receive two types of compensation: learning and wage. The model allows for general forms of production functions and incorporates varieties of sorting functions. I use the linked employee-employer data from Germany and find that workers overestimate how much they can learn from coworkers. Quantitatively, the actual learning parameters are only 12% of the perceived parameters. For example, if a worker believes her next period wage would increase by on average 10 Euros due to coworkers, it will only increase by 1.2 Euros. Namely, the learning opportunity is largely overpriced. Junior workers overpay the learning while senior workers are overpaid for the overestimated positive effects.

The model features two learning functions: one for reality and the other for belief. The correct learning function governs how knowledge evolves by learning from coworkers. Meanwhile, the perceived learning function, which may be biased from the correct one, determines the wage a worker is willing to take. I assume a competitive labor market. The compensation schedule includes a constant wage and an opportunity to learn from coworkers. Workers choose where to work by maximizing their lifetime earnings based on their belief.

The structural model only relies on team information, wages, and ages. It is based on assumptions including a competitive labor market, complete financial market, and stationarity. On the other hand, it is flexible on the firm side. The method works with general forms of complementarities across workers and production functions. Instead of working with a specific functional form, the estimation method uses the realized team composition.

I use the German matched-employee-employer data to estimate the model. The data basis is the Longitudinal Model 1975-2017 of the Linked Employer-Employee Data from the Institute for Employment Research (IAB). The data were accessed on-site at the Research Data Centre (FDZ) of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB) and via remote data execution at the FDZ. The dataset contains the complete workforce at a sample of establishments from 2008 to 2017, including characteristics like wage, age, and occupation. I use three dimensions of the data to calculate knowledge levels for each worker and identify the perceived and actual learning parameters. First, I use the wage gap in each group to back out the knowledge. Second, I use the time series of one's wage and the team composition to estimate how the actual knowledge evolves. Third, I estimate the perceived learning functions by comparing the wages of workers who have the same knowledge but work in different teams.

Operationally, identifying both the perceived and correct learning parameters depends on the same assumption, but different samples. Methodologically, it is in the same spirit as the cross-validation method. I separate the sample based on workers' knowledge. For each worker, if there is someone who has the same knowledge and age, but is in a different team, then I classify both into one of the paired groups. Otherwise, the worker belongs to the unpaired group. I use the former groups to estimate to what extent the present value of future wages can make up for the current wage gap, and use the latter group to analyze the knowledge evolvement over time.

I find that workers overestimate how much they can learn from others. The actual impact of coworkers is only 12% of one's belief. For example, if the average wage of a worker's more knowledgeable coworkers increased by 100 Euros, according to the belief, her next period wage would increase by on average 8 Euros times the share of the more knowledgeable workers in the team. But in fact, her wage only increases by the share of the more knowledgeable workers times 0.9 Euros. Namely, the learning opportunity is largely overpriced. Junior workers overpay the learning while senior workers are overpaid for the overestimated positive effects.

The results suggest that the bias in belief drives up wage inequality. In my sample, the unconditional variance of log wages is 0.21. If rather than holding the current belief, workers know the correct learning function and firms hire the same group of workers, the unconditional variance drops to 0.17. I perform a simple variance decomposition. In the data, the between-team inequality is 0.16, and the within-team inequality is 0.05. In the counterfactual case, the between-team inequality drops slightly to 0.15, and the within-team inequality decreases to 0.015. It suggests that there is positive knowledge sorting across firms, and within-team inequality is largely caused by overestimation. Quantitatively, the biased belief contributes to 71% wage inequality within teams.

**Relate Literature -** My paper provides empirical evidence on biased belief. It relates to a strand of behavior economic literature studying the judgment biases, examples include Klos et al. (2005), Rabin and Vayanos (2010), Miller and Sanjurjo (2018), Gong et al. (2020) and Andrew and Adams-Prassl (2021). Specifically, this paper finds workers are over-optimism, which is related to the overconfidence literature, for example Weinstein (1980), Carver et al. (2010), and Windschitl and Stuart (2015). This paper provides a novel way for researchers to identify **unobserved** belief by analyzing **observed** actions.

The paper is closely related to the empirical within team learning literature. The structural model of this paper is based on Jarosch et al. (2021), which estimates the peer effect empirically. The main difference is that Jarosch et al. (2021) uses within-team information, while my paper also explores the between-team characteristics. Nix (2020) also shows that there is a positive peer effect in the workplace. Broadly, this is also a part of the peer effect literature (Mas and Moretti, 2009; Cornelissen et al., 2017). I contribute to this literature by providing a method to explicitly disentangle the correct peer effect and the perceived peer effect.

This paper is also connected to the endogenous growth literature, which uses learning from others as the source of growth. Lucas Jr (2009), Lucas Jr and Moll (2014) and Perla et al. (2021) study the case in which agents randomly meet others to learn and consequentially generate growth. Other works (Jovanovic and MacDonald, 1994; Jovanovic, 2014; Perla and Tonetti, 2014; Luttmer, 2014) in growth theory also explore the growth from imitation and idea adoption. My paper contributes to this literature by providing direct evidence on sorted meetings and learning.

The remainder of this paper is organized as follows. In Section 2, I present a model where agents learn from their coworkers based on the perceived learning function, and their knowledge evolves according to the correct learning function. Section 3 reports the main results. I propose an algorithm to structurally estimate the two learning functions using the data and discuss the implications of the biased belief. Section 4 concludes.

# 2 Model

I build a model based on Jarosch et al. (2021). There is a unit mass of heterogeneous individuals in the economy with knowledge $z \in \mathcal{Z} = [0, \bar{z}]$. Individuals supply labor inelastically in a competitive labor market and consume their income. I assume that all individuals have a probability $\tau$ of dying each period. Anyone, who remains alive at age $N = 65$, leaves the market exogenously. Each period, a mass $\delta$ of new individuals enters the market at age $n_0 = 22$ with a knowledge drawn from the distribution $B_0(z)$. Agents are employed in firms and work in teams where they can learn from their coworkers. For a worker with knowledge $z$ who has coworkers $\tilde{\mathbf{z}}$, her next period knowledge level $z'$ is drawn from the distribution $G(z'|z, \tilde{\mathbf{z}})$ in reality while the distribution is $\tilde{G}(z'|z, \tilde{\mathbf{z}})$ in belief. Here I assume workers can learn from their coworkers' knowledge. In the appendix I extend the model to consider what if some part of the knowledge cannot be learnt. Financial markets are complete, so individuals maximize the expected present value of income.

Firms produce in a competitive consumption good market. To enter a market, firms pay a fixed cost $c$, and then draw a technology $a \in \mathcal{A}(a \sim A(a))$. In each period, a firm hires a team of workers $\mathbf{z}$ to produce according to its production function $F(\mathbf{z}; a)$, taking wages as given. Similar as in Jarosch et al. (2021), $F(\mathbf{z}; a)$ satisfies minimal structure requirements. Potentially, the complementarity between workers and technologies can vary across firms or periods. Therefore, teams generally are heterogeneous across firms and across time.

## 2.1 Firms

Firms hire workers in a competitive labor market and take the wage schedule as given. Use $n$ to denote the total number of workers in the team. $z_i$ is the $i$th element of $\mathbf{z}$, and $\tilde{\mathbf{z}}_{-i}$ is the set of $i$'s coworkers. The total wage $W(\mathbf{z})$ if it hires the set of workers $\mathbf{z}$ can be written as

$$W(\mathbf{z}) = \sum_{i=1}^{n} w(z_i, \tilde{\mathbf{z}}_{-i}). \tag{1}$$

A firm chooses the team of workers to maximize its profit

$$\pi(a) = \max_{\mathbf{z}} F(\mathbf{z}; a) - W(\mathbf{z}), \tag{2}$$

The optimal team setting $\mathbf{z}(a)$ satifies

$$\mathbf{z}(a) = \arg\max_{\mathbf{z}} F(\mathbf{z}; a) - W(\mathbf{z}). \tag{3}$$

## 2.2 Workers

To maintain a stationary distribution of workers, assume that the birth rate $\delta$ satisfies

$$\delta = \tau + \frac{1}{N - n_0 + 1}(1-\tau)^{N-n_0+1}, \tag{4}$$

Individuals decide where to work in each period. The payoff includes two parts: the wage $w$ and the learning opportunities. Therefore, an individual is willing to accept a lower wage if there are more attractive learning opportunities. The wage depends on both one's knowledge and the coworkers' knowledge. Individuals discount the future using a discount rate $\beta$. Because the worker leaves the market mandatorily at age $N$, the value function depends on the current age of the worker. The value function of an individual with knowledge $z$ and age $n$ is

$$V(z,n) = \max_{\tilde{\mathbf{z}} \in \tilde{\mathbf{Z}}} w(z, \tilde{\mathbf{z}}, n) + \beta \int V(z', n+1) d\tilde{G}(z'|z, \tilde{\mathbf{z}}), \tag{5}$$

where $\tilde{\boldsymbol{Z}}$ represents the set of all possible combinations of coworkers. It means that each period an individual chooses where to work to maximize her wage and her discounted future income. Based on the assumption of a competitive labor market, in equilibrium, the wage should be such that everyone is indifferent to working in any firm. Therefore, the value function does not depend on the coworker's knowledge. Specifically, when $n = N$, the value function equals the current wage

$$V(z, N) = \max_{\tilde{\mathbf{z}} \in \tilde{\mathbf{Z}}} w(z, \tilde{\mathbf{z}}, N). \tag{6}$$

This is because learning from co-workers does not matter anymore. Therefore, when $n = N$, the wage only depends on the worker's own knowledge level $z$, not the coworkers'

$$w(z, \tilde{\mathbf{z}}, N) = w(z, N), \tag{7}$$

In equilibrium, the wage schedule adjusts so that workers are indifferent across teams. The wage satisfies

$$w(z, \tilde{\mathbf{z}}, n) = V(z, n) - \beta \int V(z', n+1) d\tilde{G}(z'|z, \tilde{\mathbf{z}}), \tag{8}$$

for any $z, \tilde{\boldsymbol{z}}$ realized in the equilibrium. One implication is that for any $\tilde{\mathbf{z}}, \tilde{\mathbf{z}}'$

$$w(z, \tilde{\mathbf{z}}, n-1) - w(z, \tilde{\mathbf{z}}', n-1) = -\beta \left[ \int V(z', n) d\tilde{G}(z'|z, \tilde{\mathbf{z}}) - \int V(z', n) d\tilde{G}(z'|z, \tilde{\mathbf{z}}') \right]. \tag{9}$$

The wage difference in $(n-1)$ is used to compensate for the difference in learning opportunities. For example, two identical workers $i$ and $j$ work in different teams. If $i$ can on average learn more from working than $j$ due to the team setting, then she will receive a lower wage.

## 2.3 Labor market and Firm Entry

Define $B(z)$ the cumulative density function of $z$. For any team of workers $\mathbf{z}$, let $N(\mathbf{z}, z)$ be the number of workers whose knowledge is not larger than $z$. Labor market clearing yields

$$B(z) = m \int N(\mathbf{z}(a), z) \, dA(a), \forall z, \tag{10}$$

where $m$ is the mass of firms in the economy.

Firm entry satisfies the free entry condition

$$\int [\pi(a) - c] \, dA(a) = 0. \tag{11}$$

## 2.4 Knowledge Distribution

Define $O(\tilde{\mathbf{z}}|x) : \tilde{\mathbf{Z}} \times Z \to [0,1]$ to be the equilibrium share of workers with knowledge $x$ whose coworkers are strictly dominated by the vector $\tilde{\mathbf{z}}$. Therefore, for workers with knowledge $x$, the probability their knowledge will be less than $z$ is $\int_{\tilde{\mathbf{z}}\in\tilde{\mathbf{Z}}} G(z|x,\tilde{\mathbf{z}})\, dO(\tilde{\mathbf{z}}|x)$. The stationary distribution in equilibrium satisfies

$$B(z) = \tau B_0(z) + (1-\tau)\int_x \int_{\tilde{\mathbf{z}}\in\tilde{\mathbf{Z}}} G(z|x,\tilde{\mathbf{z}})\, dO(\tilde{\mathbf{z}}|x)\, dB(x). \tag{12}$$

## 2.5 Equilibrium

A stationary competitive equilibrium includes a wage schedule $w$, a profit function $\pi$, an employment arragement $\mathbf{z}$, a value function $V$, a mass of firms $m$, a stationary distribution $B$ and a coworker vector set $\tilde{\mathbf{Z}}$, such that: (1) $w$ and $V$ solve (2) and (5); (2) $\mathbf{z}$ solves (2); (3) the labor market clears (10) for each knowledge level; (4) $\pi$ satisfies the free entry condition (11); (5) the law of motion of the knowledge distribution 12 holds.

Next, I summarize the assumptions in the model. The goal is to ensure that the value function $V(z,n)$ strictly increases in $z$. When comparing two vectors $\mathbf{z}_1$ and $\mathbf{z}_2$, they have the same length $N$ and are ordered by the element size. $\mathbf{z}_1 > \mathbf{z}_2$ holds when each element in $\mathbf{z}_1$ is larger or equal to the corresponding elements in $\mathbf{z}_2$: $z_{1i} \geq z_{2i}, i = 1, 2, ...N$, and inequality holds for at least one element.

**Assumption 1.** *$F(\mathbf{z}; a)$ strictly increases in $\mathbf{z}$.*

This assumption means that if the knowledge of one or some workers increases, the firm's production will also increase.

**Assumption 2.** *$\tilde{G}(z'|z,\tilde{\mathbf{z}})$ and $G(z'|z,\tilde{\mathbf{z}})$ strictly decreases in $z$ and $\tilde{\mathbf{z}}$.*

Assumption 2 requires that both in belief and in reality, for two workers having the same set of coworkers, the one with higher knowledge will stochastically learn more than the other. Also, for two identical workers, both in belief and in reality, the one who works with higher knowledge coworkers will have stochastically more knowledge next period.

**Assumption 3.** *Free disposal of knowledge.*

**Assumption 4.** *Learning*

Three assumptions together lead to following results.

**Lemma 1.** *$w(z,\tilde{\mathbf{z}}_1,n) \leq w(z,\tilde{\mathbf{z}}_2,n)$, if $\tilde{\mathbf{z}}_1 > \tilde{\mathbf{z}}_2$.*

*Proof.* Assumption 2 says$\tilde{G}(z'|z,\tilde{\mathbf{z}})$ strictly decreases. The value function $V(z,n)$ is weakly increasing in $z$ according to Assumption 3. Hence, $\int V(z',n)d\tilde{G}(z'|z,\tilde{\mathbf{z}})$ increases with $\tilde{\mathbf{z}}$. $w(z,\tilde{\mathbf{z}},n)$ weakly decreases in $\tilde{\mathbf{z}}$ using Equation (9). □

**Lemma 2.** *If a firm's worker setting is* $\mathbf{z}(a)=(z_1,\tilde{\mathbf{z}})$. *Then for* $\forall z_2>z_1$, $w(z_1,\tilde{\mathbf{z}},n)<w(z_2,\tilde{\mathbf{z}},n)$ *must hold.*

*Proof.* Proof by contradiction. If there are two workers with $z_1$and $z_2$, where $z_1>z_2$ and $w(z_1,\tilde{\mathbf{z}},n)\leq w(z_2,\tilde{\mathbf{z}},n)$. It means that by hiring $z_1$, the firm can offer a lower wage to this worker. By Assumption 1, firms can produce more output. Additionally, it is less costly for firms to hire the rest workers, because of Lemma 1. Hence, firms would always want to hire $z_1$ over $z_2$. □

**Proposition 1.** $V(z,n)$ *strictly increases in* $z$.

*Proof.* Use Mathematical induction . Consider $z_1<z_2$.

First, when $n=N$, $V(z_1,N)=w(z_1,N)<w(z_2,N)=V(z_2,N)$ according to Lemma 2.

Second, assume $V(z_1,n)<V(z_2,n)$ holds for some $n$. Then

$$\begin{aligned} V(z_1,n-1) &= w(z_1,\tilde{\mathbf{z}},n-1)+\beta\int V(z',n)d\tilde{G}(z'|z_1,\tilde{\mathbf{z}}) \\ &< w(z_2,\tilde{\mathbf{z}},n-1)+\beta\int V(z',n)d\tilde{G}(z'|z_2,\tilde{\mathbf{z}}) \\ &= V(z_2,n-1) \end{aligned} \tag{13}$$

Hence, $V(z,n)$ strictly increases in $z$. □

# 3 Structural Estimation

Based on the model, this section structurally estimates the perceived and actual learning functions $\tilde{G}(\cdot)$ and $G(\cdot)$. The data basis is the Longitudinal Model 1975 – 2017 of the Linked Employer-Employee Data from the IAB. The data were accessed on-site at the Research Data Centre (FDZ) of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB) and via remote data execution at the FDZ. The dataset contains the complete workforce at a sample of establishments from 2008 to 2017.

## 3.1 Data

This section briefly describes the key dimensions of the dataset. The establishment part includes all establishments that are surveyed by an annually conducted survey at least once from 2009

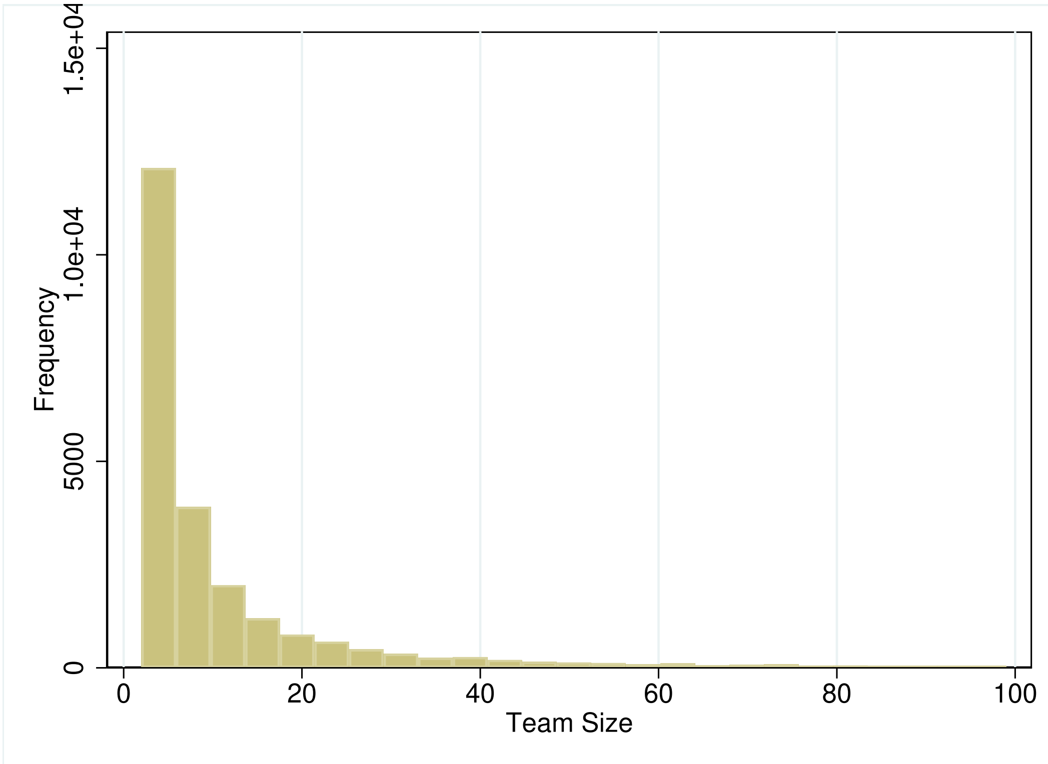


Figure 1: Team Size Distribution

Note: The figure plots the team size distribution in the year 2009 for teams with 2-99 workers.

to 2016. The individual part includes workers who were employed in any one of the sample establishments for at least one day during 2008 - 2017. For each individual, it also contains the employment biographies from 1975 - 2017. The data include worker-level information on the establishment, occupation, wage, and individual characteristics like age, gender, and education. I define working groups following Jarosch et al. (2021): a group is defined as a set with at least two workers in the same establishment and occupation in a given year.

Figure 1 reports the team size distribution in 2009. The figure truncated at 100 workers per team. The sample contains 24,118 teams. The size distribution is skewed, with a median of 5 people in contrast to an average of 23.

Figure 2 shows the age distribution of all workers in teams. The average age is 43. The 10th, 50th, and 90th percentiles are 28, 44, and 56, respectively.

Figure 3 reports the distribution of the average daily wages during the year 2009 in 2015 Euros. The mass-point observations are the top-coding of the wage data. [1]

This paper builds upon the wage difference within teams. Figure 4 shows the distribution of the gap between the log wage of a worker and her coworkers. The mean is -0.015, the quartiles

[1]Following Jarosch et al. (2021), I treat the top-coded observations as the actual wages and do not correct for top-coding. In total, 10.6% of observations are top-coded.

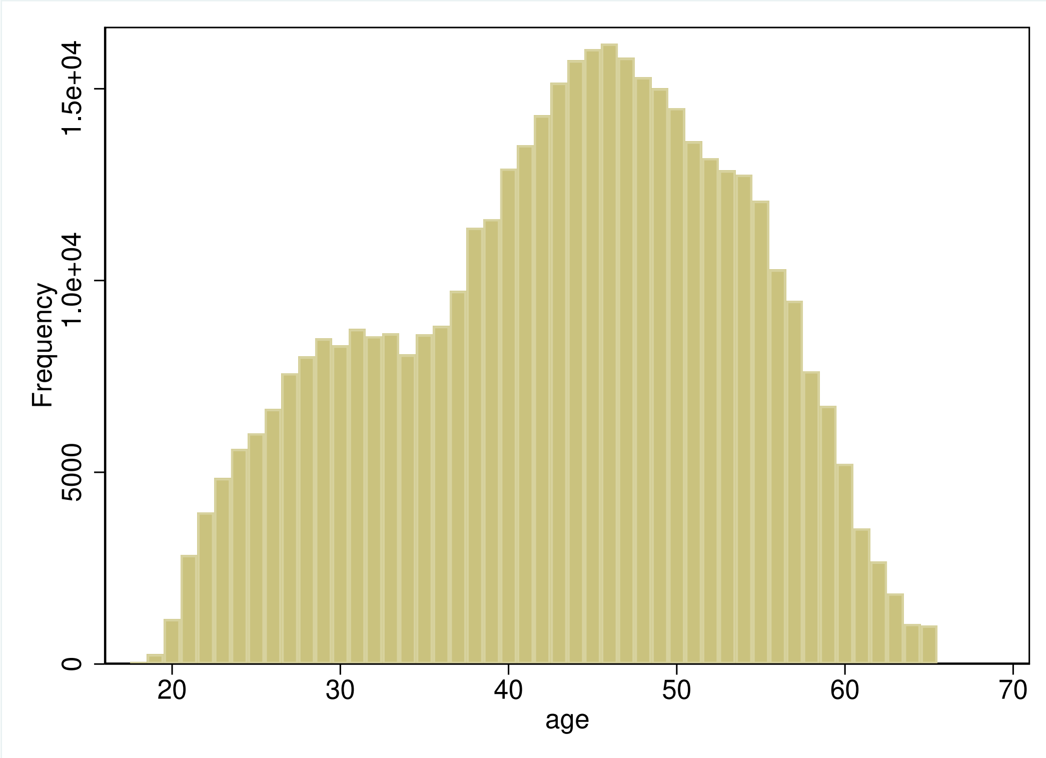


Figure 2: Age Distribution

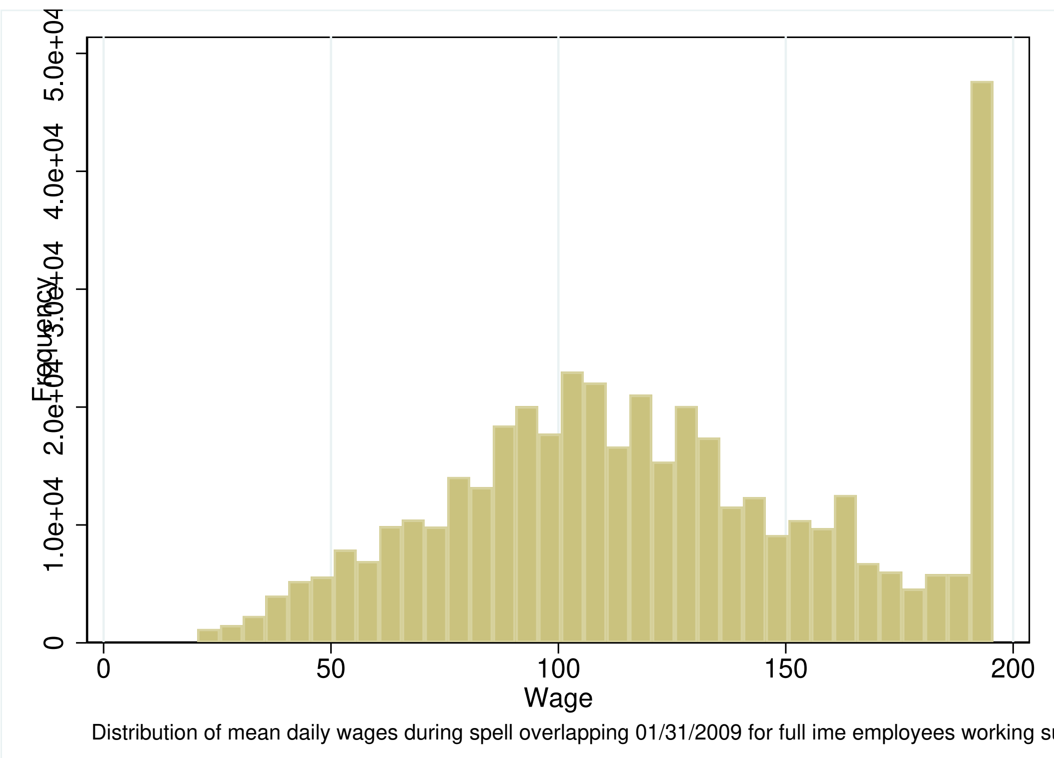


Figure 3: Wage Distribution

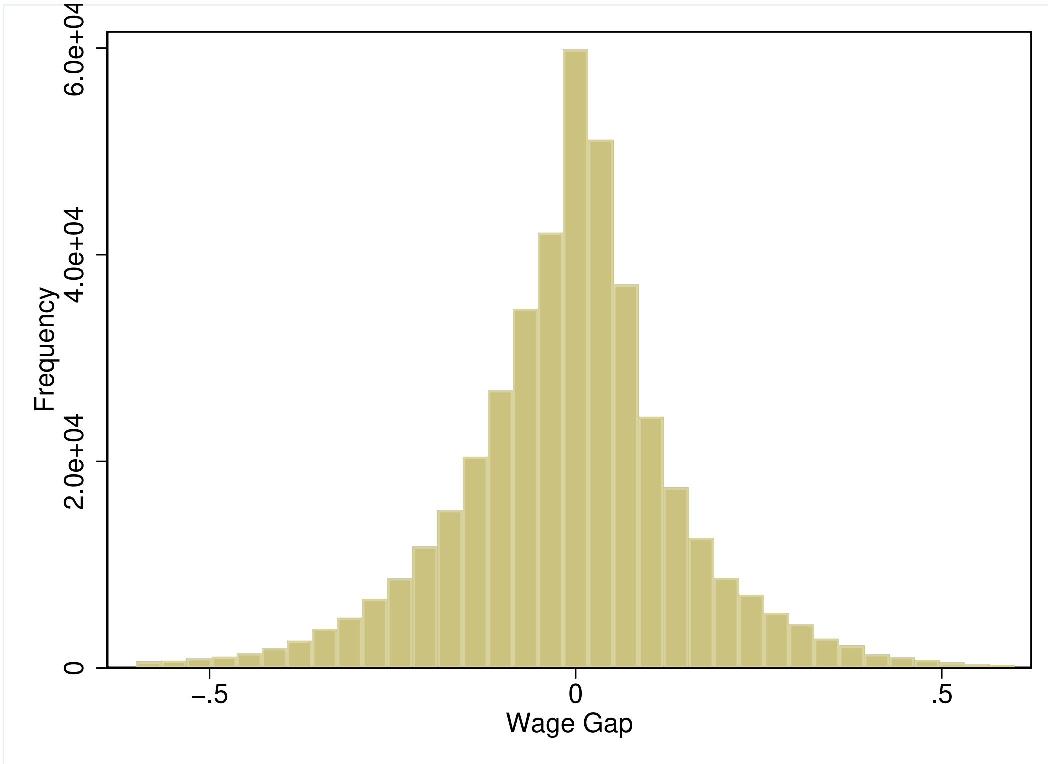


Figure 4: Wage Gap Distribution

are -0.09, 0.00, and 0.07.

## 3.2 Identifying Perceived And Actual Learning Parameters

I use three dimensions of the data to identify the perceived and actual learning parameters, as well as calculate knowledge levels for each worker. First, I use the wage gap in each group to back out the knowledge. Second, I use the time series of one's wage and the team composition to estimate how the actual knowledge evolves. Third, I estimate the perceived learning functions by comparing the wages of workers who have the same knowledge level but work in different teams.

The identification only relies on the worker's Bellman equation

$$V(z,n) = w(z, \tilde{\mathbf{z}}, n) + \beta \mathbb{E}(V(z', n+1)|z, \tilde{z}). \tag{14}$$

The method does not impose any restrictions on firms' production in addition to the previous assumptions, nor on the set of active firms – it simply uses all observed firms.

Because $z$ does not have a natural cardinality, I choose a convenient one. If $w(z,n)$ is the wage for workers who works along in the equilibrium, I choose a cardinality of $z$ so that

$w(z,n) = z$. Therefore, Equation 14 can be written as

$$V(z,n) = z + \beta \int V(z', n+1) d\tilde{G}(z'|z), \tag{15}$$

with a boundary condition

$$V(z, N) = z. \tag{16}$$

Worker $i$ next period's knowledge includes two parts

$$z_i' = \mathbb{E}\left(z_i' | z_i, \tilde{\mathbf{z}}_{-i}\right) + \varepsilon_i, \tag{17}$$

a conditional expectation and an unexpected shock. All moment conditions are based on $\mathbb{E}\left(\varepsilon_i | z_j\right), \forall i, j$, which is implied by the assumption that the shock is independent of workers' knowledge.

I choose the functional forms for the two learning functions based on Jarosch et al. (2021). The actual learning function takes the following parametric form:

$$\mathbb{E}\left(z_i' | z_i, \tilde{\mathbf{z}}_{-i}\right) = \frac{1}{I-1} \sum_{j \neq i} z_i \Theta\left(\frac{z_j}{z_i}\right), \tag{18}$$

and the perceived learning function is

$$\tilde{\mathbb{E}}\left(z_i' | z_i, \tilde{\mathbf{z}}_{-i}\right) = \frac{1}{I-1} \sum_{j \neq i} z_i \tilde{\Theta}\left(\frac{z_j}{z_i}\right). \tag{19}$$

$I$ is the team size and both $\Theta(\cdot)$ and $\tilde{\Theta}$ are weakly increasing function. Two learning functions share the same functional form but different parameters:

$$\Theta\left(\frac{z_j}{z_i}\right) = \begin{cases} 1 + \theta_0 + \theta^+ \left(\frac{z_j}{z_i} - 1\right) & \text{if } \frac{z_j}{z_i} > 1 \\ 1 + \theta_0 + \theta^- \left(\frac{z_j}{z_i} - 1\right) & \text{if } \frac{z_j}{z_i} \leq 1 \end{cases} \tag{20}$$

and

$$\tilde{\Theta}\left(\frac{z_j}{z_i}\right) = \begin{cases} 1 + \tilde{\theta}_0 + \tilde{\theta}^+ \left(\frac{z_j}{z_i} - 1\right) & \text{if } \frac{z_j}{z_i} > 1 \\ 1 + \tilde{\theta}_0 + \tilde{\theta}^- \left(\frac{z_j}{z_i} - 1\right) & \text{if } \frac{z_j}{z_i} \leq 1 \end{cases} \tag{21}$$

Operationally, this paper estimates a case where $\theta_0 = \tilde{\theta}_0, \theta^+ = \gamma\tilde{\theta}^+, \theta^- = \gamma\tilde{\theta}^-$. Intuitively, it means workers know how knowledge evolves without learning from coworkers. They may have a misperception about how much they are affected by coworkers though, due to reasons such as bias when estimating others' knowledge levels.

The value function yields based on the chosen functional form

$$V(z_i, n_i) = z_i + \beta V((1+\theta_0) z_i, n_i + 1). \tag{22}$$

Use

$$V(z_i, N) = z_i, \tag{23}$$

and solve backwards gives

$$\begin{aligned} V(z_i, n_i) &= \sum_{t=0}^{N-n_i} [\beta(1+\theta_0)]^t z_i \\ &= \tilde{\beta}_i z_i \end{aligned} \tag{24}$$

where $\tilde{\beta}_i = \frac{1-\left[\beta\left(1+\tilde{\theta}_0\right)\right]^{N-n_i+1}}{1-\beta\left(1+\tilde{\theta}_0\right)}$. The Bellman equation is

$$\tilde{\beta}_i z_i = w_i + \tilde{\mathbb{E}}\left(z_i' | z_i, \tilde{\mathbf{z}}_{-i}\right). \tag{25}$$

There is one equation for each worker. Given perceived learning parameters $\left\{\tilde{\theta}_0, \tilde{\theta}^+, \tilde{\theta}^-\right\}$, the knowledge level $z_i$ is a function of wages and ages.

I estimate the perceived parameter $\gamma$ by comparing wage gaps between workers with the same knowledge level and same age. If workers $i$ and $j$ have the same knowledge level and the same age, then from Equation (9) their wages satisfy:

$$\tilde{\beta}_i\left(z_i' - z_j'\right) = -\gamma\left(w\left(z_i, \tilde{\mathbf{z}}_{-\mathbf{i}}, n_i\right) - w\left(z_j, \tilde{\mathbf{z}}_{-j}, n_j\right)\right) - \frac{\tilde{\beta}_i}{\gamma}\left(\varepsilon_j - \varepsilon_i\right). \tag{26}$$

The moment conditions are

$$\mathbb{E}\left(z_j \varepsilon_i | \exists k, z_k = z_i\right) = 0 \tag{27}$$

For each age $n_i$ and corresponding $\tilde{\beta}_i$, there is one moment condition. Operationally, I use the moment conditions for $n = \{20, 21, ...64\}$. Together, the 45 moments identify $\gamma$.

I estimate the actual learning parameters by tracking the evolvement of knowledge

$$z_i' = (1+\theta_0) z_i + \frac{1}{I-1}\left\{\theta^- \sum_{z_j \le z_i} (z_j - z_i) + \theta^+ \sum_{z_j > z_i} (z_j - z_i)\right\} + \varepsilon_i. \tag{28}$$

The parameters can be estimated $\left\{\tilde{\theta}_0, \tilde{\theta}^+, \tilde{\theta}^-\right\}$ by running a linear regression.

The algorithm is described below:

1. For every $\gamma$ in a reasonable range $[\gamma_-, \gamma_+]$, guess the perceived learning parameters $\left\{\tilde{\theta}_0, \tilde{\theta}^+, \tilde{\theta}^-\right\}^{Guess}$.

2. Given the $\gamma$ and $\left\{\tilde{\theta}_0, \tilde{\theta}^+, \tilde{\theta}^-\right\}^{Guess}$, I can back out the knowledge $z$ for all workers using Equation (25).

3. Pair workers according to knowledge and age: two workers are in a group if both knowledge and ages are the same. Workers who are paired are grouped into subgroups by age, and the rest are grouped together.

4. Use the unpaired group to run regression (28). Update the learning parameters $\left\{\tilde{\theta}_0, \tilde{\theta}^+, \tilde{\theta}^-\right\}^{Guess}$ accroding to the regression.

5. Repeat step 2 to 4 until convergence.

6. Use the paired groups to estimate the moment conditions from equation system 26.

7. Find the $\gamma$ that minimizes the squared difference between data and model.

Identifying both $\left\{\tilde{\theta}_0, \tilde{\theta}^+, \tilde{\theta}^-\right\}$ and $\gamma$ depends on the assumption that shocks $\varepsilon$ are independent with knowledge $z$. The key is to use two separate and representative samples to estimate the two sets of parameters. Methodologically, it is in the same spirit as the cross-validation method. In practice, I use the 45 paired samples to estimate $\gamma$, that is, to what extent the present value of future wages can make up for the current wage gap. I use the unpaired sample to analyze the knowledge evolution over time and to estimate the learning parameters.

## 3.3 Estimation Results

Using the estimation method developed to estimate both the perceived learning function $\left\{\tilde{\theta}_0, \tilde{\theta}^+, \tilde{\theta}^-\right\}$ and the perceived parameter $\gamma$. The results are reported in Table (1). The perceived learning parameters are reported in panel A. $\tilde{\theta}_0$ is positive, which represents that workers accumulate knowledge during work. This is similar to the common sense of learning-by-doing. The perceived learning from more and less knowledgeable workers is 0.08 and 0.05, respectively. It means that workers think that they are affected asymmetrically by their coworkers: workers are more heavily impacted by more knowledgeable coworkers than the rest. The numbers can be interpreted in terms of expected wage changes. Consider a case in which a worker's more knowledgeable coworkers become even better. For example, if the average wage of a worker's more knowledgeable coworkers increase by 100 Euros, according to the belief, her next period wage will increase by on average 8 Euros times the share of the more knowledgeable coworkers in the team. On the other hand, if the more knowledgeable coworkers stay the same, while the less knowledgeable coworkers become better with a 100 Euros wage increases, then her next period wage will go up by on average 5 Euros times the share of the more knowledgeable workers in the team. This estimation is consistent with the results in Jarosch et al. (2021). They also find workers are more affected by more knowledgeable coworkers.

The key finding is that the difference between real life and perception is huge, with $\gamma = 0.12$. It means that the actual impact of coworkers is only 12% of one's belief. Therefore, in the previous example, when the average wage of a worker's more knowledgeable coworkers increase by 100 Euros, her next period wage will increase by on average the share of the more knowledgeable coworkers in the team times 0.9 Euros, instead of 8 Euros. If the more knowledgeable coworkers stay the same, while the less knowledgeable coworkers become better with a 100 Euros wage increases, then her next period wage will go up by on average the share of the more knowledgeable workers in the team times 0.6 Euros, rather than 5 Euros. It means that workers overestimate how much they can learn from coworkers. Namely, the learning opportunity is largely overpriced. Junior workers overpay the learning while senior workers are overpaid for the overestimated positive effects.

Table 1: Estimation Results

**Panel A: GMM Estimation**

| | $\tilde{\theta}_0$ | $\tilde{\theta}^+$ | $\tilde{\theta}^-$ | $\gamma$ | obs |
|---|---|---|---|---|---|
| Estimation | 0.01688 | 0.07826 | 0.05000 | 0.120 | 4188203 |
| | (0.000001) | (0.00003) | (0.00003) | (0.005) | |

**Panel B: The Actual Learning Parameters Based on the Estimation**

| | $\theta_0$ | $\theta^+$ | $\theta^-$ |
|---|---|---|---|
| Results | 0.01688 | 0.00939 | 0.00600 |

[1] Note: the standard deviation is reported in the parenthesis. Belief is estimated using GMM. The reality is inferred based on the estimation.

The estimation can be extended to incorporate other form of learning functions. For example, the assumption $\theta_0 = \tilde{\theta}_0, \theta^+ = \gamma\tilde{\theta}^+, \theta^- = \gamma\tilde{\theta}^-$ can be relaxed. The same algorithm will still work.

## 3.4 Inequality

This section considers what if the belief were correct, that is, $\theta_0 = \tilde{\theta}_0, \theta^+ = \tilde{\theta}^+, \theta^- = \tilde{\theta}^-$. According to the structural model, the learning opportunities contribute to the wage difference

Table 2: Inequality Change When Perception Is Correct

| | Benchmark | Correct Belief |
|---|---|---|
| Wage Inequality | 0.210 | 0.167 |
| Between Teams Wage Inequality | 0.158 | 0.152 |
| Within Team Wage Inequality | 0.052 | 0.015 |

Note: the table reports the change in inequality when there is no bias in belief. The inequality is measured by unconditional variance of log wages

and consequentially income inequality. The overestimated learning opportunity, therefore, affects the income inequality. Assume that suddenly people realize that their belief is biased and know the correct learning function. I focus on a special case where firms are willing to adjust the wages according to the new belief without changing team compositions. As a result, the wages of workers with more knowledge go down, since they were overpaid before. Meanwhile, less knowledgeable workers receive a wage raise.

The inequality levels are shown in Table (2). The first column shows the total wage inequality in the data, which is measured by the variance of log wages. The second column reports the wage inequality if there is no bias in belief. The total wage inequality drops from 0.21 to 0.17. To explore the mechanism behind, I calculate a simple inequality decomposition and find that the changes are mainly from the within-team inequality. It decreases by more than 70%: from 0.052 to 0.015. This result implies that the overestimation contributes significantly to the within-team wage difference, where more knowledgeable workers are paid even more. Moreover, the results show that the between team inequality almost remains unchanged: it is 0.158 in the data and 0.152 under the correct belief. This suggests that there is positive sorting between firms.

# 4 Conclusion

This paper shows evidence suggesting people overestimate the magnitude of learning from coworkers. As a result, senior workers are overpaid for their overestimated impact on others, while junior workers are overcharged for the overstated learning opportunities they have by working with experienced coworkers. This suggests that the biased belief may lead to an increase in within-team wage inequality.

This paper points to a plausible way to reduce inequality: to narrow the discrepancy between the perceived and correct learning function. It also provides suggestive evidence on the explanation of wage inequality. The between-firm inequality is largely due to knowledge differences while the within-firm inequality is mainly driven by learning opportunities. There are multiple paths for future research. One is to think about optimal policies to reduce inequality conditional on belief bias. Another is to explore the heterogeneity in bias and how to reduce biases.

# References


**Andrew, Alison, and Abi Adams-Prassl.** 2021. "Revealed beliefs and the marriage market return to education."Technical report, Working paper.

**Carver, Charles S, Michael F Scheier, and Suzanne C Segerstrom.** 2010. "Optimism." *Clinical psychology review* 30 (7): 879–889.

**Cornelissen, Thomas, Christian Dustmann, and Uta Schönberg.** 2017. "Peer Effects in the Workplace." *American Economic Review* 107 (2): 425–56.

**Gong, Yifan, Ralph Stinebrickner, and Todd Stinebrickner.** 2020. "Marriage, children, and labor supply: Beliefs and outcomes." *Journal of Econometrics*.

**Jarosch, Gregor, Ezra Oberfield, and Esteban Rossi-Hansberg.** 2021. "Learning From Coworkers." *Econometrica* 89 (2): 647–676. 10.3982/ecta16915.

**Jovanovic, Boyan.** 2014. "Misallocation And Growth." *American Economic Review* 104 (4): 1149–71.

**Jovanovic, Boyan, and Glenn M MacDonald.** 1994. "The life cycle of a competitive industry." *Journal of Political Economy* 102 (2): 322–347.

**Klos, Alexander, Elke U Weber, and Martin Weber.** 2005. "Investment decisions and time horizon: Risk perception and risk behavior in repeated gambles." *Management Science* 51 (12): 1777–1790.

**Lucas Jr, Robert E.** 2009. "Ideas and growth." *Economica* 76 (301): 1–19.

**Lucas Jr, Robert E, and Benjamin Moll.** 2014. "Knowledge Growth and the Allocation of Time." *Journal of Political Economy* 122 (1): 1–51.

**Luttmer, Erzo GJ.** 2014. "Knowledge diffusion, growth, and inequality." *University of Minnesota, Working Paper (forthcoming)*.

**Mas, Alexandre, and Enrico Moretti.** 2009. "Peers at Work." *American Economic Review* 99 (1): 112–45.

**Miller, Joshua B, and Adam Sanjurjo.** 2018. "Surprised by the Hot Hand Fallacy? A Truth in the Law of Small Numbers." *Econometrica* 86 (6): 2019–2047.

**Nix, Emily.** 2020. "Learning spillovers in the firm."Technical report, Working Paper.

**Perla, Jesse, and Christopher Tonetti.** 2014. "Equilibrium imitation and growth." *Journal of Political Economy* 122 (1): 52–76.

**Perla, Jesse, Christopher Tonetti, and Michael E. Waugh.** 2021. "Equilibrium Technology Diffusion, Trade, and Growth." *American Economic Review* 111 (1): 73–128. 10.1257/aer.20151645.

**Rabin, Matthew, and Dimitri Vayanos.** 2010. "The Gambler's and Hot-hand Fallacies: Theory and Applications." *The Review of Economic Studies* 77 (2): 730–778.

**Weinstein, Neil D.** 1980. "Unrealistic optimism about future life events.." *Journal of personality and social psychology* 39 (5): 806.

**Windschitl, Paul D, and Jillian O'Rourke Stuart.** 2015. "Optimism biases: Types and causes." *The Wiley Blackwell handbook of judgment and decision making* 2 431–455.

# A Multi-dimension Knowledge

In the baseline model, I assume that workers can learn from their coworkers' knowledge. One concern is that the knowledge, as a multi-dimension aggregation, may include components that one cannot learn. This section studies the impact of non-transmissible knowledge. Assume that the knowledge $z = x + y$ includes two parts: transmissible $x$ and non-transmissible $y$. A worker's next period non-transmissible knowledge level stays the same, and the transmissible knowledge level $x'$ is drawn from the distribution $\hat{G}(x'|x, \tilde{\mathbf{x}})$. Assume the actual learning function takes the same functional form as (18)

$$\mathbb{E}(x_i'|x_i, \tilde{\mathbf{x}}_{-i}) = (1+\theta_0)x_i + \frac{1}{I-1}\left(\theta^+ \sum_{x_j > x_i} (x_j - x_i) + \theta^- \sum_{x_j < x_i} (x_j - x_i)\right) \tag{29}$$

Thus, the knowledge $z$ follows

$$\mathbb{E}(z_i'|z_i, \tilde{\mathbf{x}}_{-i}) = y_i + (1+\theta_0)x_i + \frac{1}{I-1}\left(\theta^+ \sum_{x_j > x_i} (x_j - x_i) + \theta^- \sum_{x_j < x_i} (x_j - x_i)\right) \tag{30}$$

The same functional form applies to the perceived learning function $\hat{\tilde{G}}(x'|x, \tilde{\mathbf{x}})$. The wage function is

$$w\left(x_i, y_i, \tilde{\mathbf{x}}_{-\mathbf{i}}, n_i\right) = x_i + y_i - \frac{\tilde{\beta}_i}{I-1}\left(\theta^+ \sum_{x_j > x_i}\left(x_j - x_i\right) + \theta^- \sum_{x_j < x_i}\left(x_j - x_i\right)\right) \tag{31}$$

In the benchmark model, a worker's next period knowledge level $z'$ is assumed to be drawn from the distribution $G(z'|z, \tilde{\mathbf{z}})$. The wage function is

$$w\left(z_i, \tilde{\mathbf{z}}_{-\mathbf{i}}, n_i\right) = z_i - \frac{\tilde{\beta}_i}{I-1}\left(\theta^+ \sum_{z_j > z_i}\left(z_j - z_i\right) + \theta^- \sum_{z_j < z_i}\left(z_j - z_i\right)\right) \tag{32}$$

Potentially, the misspecification means workers are only affected by partial of their coworkers' knowledge but the model treats it as if workers are affected by their coworkers' aggregate knowledge. To understand how the misspecification affects the result, I consider two specific functional forms: the non-transmissible part is linear or independent of knowledge. In both cases, the belief bias is not affected by the learning function misspecification.

## A.1 Case 1: the non-transmissible is linear in knowledge

Assume that $y = \alpha x$. The wage is

$$w\left(x_i, y_i, \tilde{\mathbf{x}}_{-\mathbf{i}}, n_i\right) = (1+\alpha)x_i - \frac{\tilde{\beta}_i}{I-1}\left(\theta^+ \sum_{x_j > x_i}\left(x_j - x_i\right) + \theta^- \sum_{x_j < x_i}\left(x_j - x_i\right)\right). \tag{33}$$

which can be written as

$$w\left(z_i, \tilde{\mathbf{z}}_{-\mathbf{i}}, n_i\right) = z_i - \frac{\tilde{\beta}_i}{I-1}\left(\frac{\theta^+}{1+\alpha} \sum_{z_j > z_i}\left(z_j - z_i\right) + \frac{\theta^-}{1+\alpha} \sum_{z_j < z_i}\left(z_j - z_i\right)\right). \tag{34}$$

If the model is estimated using the misspecified model in Equation (32), it is the same as using equation (34). Therefore, the estimated knowledge level is $z$, the aggregate knowledge level. The estimated learning parameters $\hat{\theta^+}$ and $\hat{\theta^-}$ satisfy that $\hat{\theta}^+ = \frac{\theta^+}{1+\alpha}$, $\hat{\theta}^- = \frac{\theta^-}{1+\alpha}$. The same bias applies to the perceived learning parameters $\hat{\tilde{\theta}}^+ = \frac{\tilde{\theta}^+}{1+\alpha}$, $\hat{\tilde{\theta}}^- = \frac{\tilde{\theta}^-}{1+\alpha}$. As a result, the relationship between $\theta, \theta$ and $\tilde{\theta}^+, \tilde{\theta}^-$ – the estimation of $\gamma$ is not affected by the misspecification.

## A.2 Case 2: the non-transmissible is independent with knowledge

Assume that $y_i \perp z_j$, $\forall i, j$. The wage is

$$w\left(x_i, y_i, \tilde{\mathbf{x}}_{-\mathbf{i}}, n_i\right) = x_i + y_i - \frac{\tilde{\beta}_i}{I-1}\left(\theta^{+}\sum_{x_j > x_i}\left(x_j - x_i\right) + \theta^{-}\sum_{x_j < x_i}\left(x_j - x_i\right)\right). \tag{35}$$

which can be written as

$$w\left(z_i, \tilde{\mathbf{z}}_{-\mathbf{i}}, n_i\right) = z_i - \frac{\tilde{\beta}_i}{I-1}\left(\theta^{+}\sum_{z_j > z_i}\left(z_j - z_i\right) + \theta^{-}\sum_{z_j < z_i}\left(z_j - z_i\right)\right) + u_i. \tag{36}$$

where $u_i = \frac{\tilde{\beta}_i}{I-1}\left(\theta^{+}\sum_{z_j > z_i}\left(y_j - y_i\right) + \theta^{-}\sum_{z_j < z_i}\left(y_j - y_i\right)\right)$, and $u_i \perp x_j, \forall i, j$. Therefore, the estimated knowledge level is $\hat{z}_i = z_i + \eta_i$, $z_i \perp \eta_i$. The regression results are not affected by the misspecification; namely, the estimation of $\gamma$ is not affected by the misspecification.