The Economics of Human Capital Investment, Rents and Wage Inequality

Akbar Ullah

Division of Economics, School of Business
University of Leicester

A thesis submitted for the Degree of Doctor of Philosophy at the University of Leicester.

2019

To my Mother and Sweety

The Economics of Human Capital Investment, Rents and Wage Inequality

by

Akbar Ullah

This thesis examines three different, self-contained topics in the field of labour economics. The first chapter presents evidence which shows that firms are charging rent on training, capital and R&D investment, but they do not charge such a rent on working hours. It shows that under standard production functions, this evidence is inconsistent with models of human capital where wage setting takes place through bargaining and/or the worker's quit decision is exogenous. Then, it develops a model where the worker has no bargaining power and makes optimal quit decision to better explain the evidence. In such a model, the firm optimal wage strategy is to charge rent on only those factors which are under firm's direct control such as workers training. The model also contributes to the literature that tries to explain firms' investment in workers transferable skills.

The second chapter addresses the findings that workers who have high pre-job schooling also get more training during jobs. It develops a model of investment in human capital which shows that training can increase in pre-job schooling only if schooling increases efficiency of training in human capital production or affects worker's preferences. But except the positive efficiency effects, all the possible preferences effects of schooling are on their own not enough to generate increasing wage function in schooling. When schooling has a positive role in human capital production, then human capital, consumption and wages are strictly increasing in pre-job schooling. The model generated net of training cost wage distribution can at least partly explain the US workers' earnings distribution by education categories. Unlike standard labour-leisure choice models a novel and empirically testable prediction of the model is that, although highly qualified workers work for less time, they spend this time in learning rather than in leisure.

Chapter three addresses the conflicting evidence about the wage return of training. Longitudinal panel data based studies during the 1990s and the 2000s give very high wage return of training. On the other hand, studies in the 2010s that are based on randomised experiments estimate insignificant return from training. Using the recently developed heterogeneity robust techniques, this chapter shows that training's immediate wage effects are low on average but it has high and significant long term wage effects. The chapter concludes that the experiment based studies suffer from small sample size and short time span as in these studies only immediate post-training period data is considered for the analysis.

Acknowledgements

I would like to thank my supervisors Dr. Richard M. H. Suen and Dr. Dimitrios Varvarigos for their help, support, encouragement, and of course their patience. I learned a lot from their expertise and empathy. At this moment I cannot forget my ex-supervisor professor Martin K. Jensen who I found a really kind person besides his knowledge.

My special thanks and love goes to my wife Sara Jabeen whose continuous care and support went a long way in helping to complete my studies. She was a wonderful support at the time when I would dream equations in my sleep due to depression and anxiety in my early days of PhD research.

I would also like to thank the Division of Economics and the Business school for their financial support. Without the Graduate Teaching Assistantship I would not have been able to complete my studies. I would particularly like to thank Dr. Subir Bose and Ms Samantha Hill for their support. Sam you are a wonderful administrator I have ever witnessed in my career. I would also like to thank all the staff members in the Economics Division and PhD colleagues in the University whose support has enabled me to complete this journey.

Declaration

I declare that:

All chapters are sole authored,

A different version of chapter's 2 model has been presented at the RES PhD Meeting and Job Market 2018, held at the Westminster Business School, London,

Chapter 3 has been presented at the International PhD Meeting in Economics, 2018, University of Macedonia, Thessaloniki Greece.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This dissertation addresses three different topics which arise from theoretical and empirical research in the labour market. Chapter two focuses on the firm's behavior in terms of rent charging. It reports existing empirical studies that estimate productivity and wage return of job training and finds that firms consistently charge rent on training, capital and R&D investment. However, the findings on whether the firm charges rent on working hours are ambiguous at best in these studies. The chapter then uses Belgium firms panel and estimates that firms are not charging rent on working hours even though they charge rents on capital and training. To understand this, one needs to study the type of wage setting behavior that can lead to such a finding. The chapter shows that the existing non-competitive models which rely on wage bargaining and/or on the assumption that workers quit decisions are exogenous cannot explain such a finding under the standard assumption that different inputs are complements in the production function.

To address the aforementioned finding, the chapter develops a model where the firm invests in training and the worker makes endogenous quit decisions but has no bargaining power in terms of wage setting. It shows that in such a setting the firm does not charge rent on working hours or on any other factor that is decided by the worker in the model. On the other hand, the firm charges rent on factors which the firm controls such as training. Though the above mentioned empirical investigation may have weaknesses, this study at least highlights the need for further research in this area. Future studies can use more detailed matched employer-employee data to reconfirm the above evidence, and check what type of non-competitive wage setting processes can explain the data better. This chapter also contributes to the existing literature which tries to address the question of why firms invest in workers' general transferable skills even if workers are free to leave. Besides relaxing assumptions of the earlier models, it shows that a model where the firm invests in both general and specific skills can explain the empirical evidence relatively better than the asymmetric information based explanations.

The third chapter focuses on the empirical evidence, which shows that workers with high pre-job qualifications get more job training. Firstly, this chapter tries to answer why already qualified individuals would like to get more training during the job. Secondly, it tries to figure out the long term implication of a positive qualification-training association for wage inequality among workers with different pre-job qualifications. It develops a heterogeneous agents macroeconomics model of human capital accumulation and growth to answer the above questions. In the model, the worker distributes time between work, training and leisure and pays the direct cost of training besides the forgone income when the worker is on training instead in the routine work.

The chapter shows that the individual's time investment in training can increase in their pre-job schooling if schooling either improves the efficiency of training time in human capital accumulation or affects worker's preferences. But all the possible preferences effects of schooling are on their own not enough to generate increasing human capital and wage functions in schooling. On the other hand, in the case of direct human capital effects of schooling, net of investment costs wage function, consumption and human capital increase in pre-job schooling. The generated wage function can at least partly match the US workers' median earning distribution by schooling categories. A novel and interesting result from this model is its prediction that the leisure time is not decreasing in schooling. Unlike the standard labour-leisure choice models which predict that work time decreases and resultantly leisure time must increase in income, this model predicts that the more qualified individuals spend the time they get from working less in learning rather than in leisure. This prediction is important for wage implication of differences in schooling and can be tested from household surveys in the future by regressing leisure time on qualifications.

The fourth chapter is about the estimates of wage return from training. Studies that are based on panel data from household surveys during the 1990s and the 2000s give wage return from one week of training which, in many cases, is equivalent to a wage return from one year of schooling. But these studies are put in doubt on two fronts later on. First, randomised experiments based studies in the 2010s estimate an insignificant and low wage return from training. Second, many recent papers found the fixed-effects econometric techniques, which are also dominantly used in the wage return estimation, not robust to heterogeneities in the multi-periods setting. This chapter develops two sets of panel data from the British Household Panel Surveys. It then re-estimates the wage return to training using the recently developed multi-periods weighted difference-in-difference techniques to see if there is any over estimation of wage return from training in earlier studies. The results suggest that though the standard fixed-effects methods used in earlier studies are not robust to heterogeneity in treatment effects, the return from training is still positive, high and significant. The return estimates of training from the heterogeneity robust techniques are low during or in the immediate post-training period, but it increases in later periods. Thus, training seems to have dynamic effects on the wage return for many years. The chapter concludes that the low and insignificant return estimates from studies in the 2010s are due to the fact that these studies use only single period immediate post-training data to estimate training return.

Before delving into the details, below are some statistics on the status and importance of training in the first instance. The earning gap between the median college educated and median high school educated US male workers working full-time in year-round jobs rose from \$17, 411 in the year 1979 to \$34,969 in 2012 (Autor, 2014). For female workers, the corresponding gap rose from \$12,887 to \$23,280 between 1979 and 2012. Similar trends exist in all OECD countries from Autor's (2014) calculations and a number of other studies (OECD, 2010, He, 2012). To explain this, the literature mostly focuses on skill biased technical changes over the years. It is true that such demand side changes are playing important role in the increasing wage gap, but the role of skills accumulation cannot be ignored. As mentioned above, not only highly qualified individuals are found getting more training during jobs but such trends seem to increase over time. For example, the British Skills and Employment Survey 2006 shows that the proportion of people strongly agreeing to the statement that 'my job requires that I keep learning new things' has increased from 26% in 1986 to 35% in 2006 (Felstead et al., 2007). Similarly, the use of automated or computerised equipment at jobs has increased substantially according to this survey. According to the Pew Research Center (2016) report, 54% of the adults in the US labour market think that getting training to develop new skills to keep up with the changing workplace is essential throughout their life. Even among those employed workers who think that their skills and education is enough to go ahead in their career, 47% think that ongoing training is essential throughout their career. For both the US and UK, the latest waves of these surveys show that around two-thirds of the workers think that most of the skills needed at work are learned on the job rather than in school or university (Pew Research Center, 2016, Henseke et al., 2018).

Regarding actual participation in training, the Pew Research Center (2016) survey shows that 45% of the US workers have taken a class or got training over the last 12 months in order to get, maintain or improve skills. According to the 2012 wave of Skills and Employment Survey, the proportion of UK's workers who participated in a more than 10 days training is 34%. This survey also reports a significant increase in self-teaching and correspondence/internet training courses, leading to an overall participation rate of 68% in the year 2012 (Felstead et al., 2013). On the expenditure on training, the most detailed and biggest survey covering training is the Employer Skills Survey of UK covering 87,430 employers in 2017. According to this, 66% of the establishments trained their staff over the last 12 months incurring training expenditure of £44.2 billion (Winterbotham et al., 2018). 62% of the staff were trained and the expenditure per trained employee is £2,470. According to the Association for Talent Development report for US[1], the annual training expenditure per employee in the US is $1,296 in the year 2018. Training expenditure and participation in training is even higher than this in Japan and some European countries. For example, the German apprenticeship is a classic example of on-the-job training and more than 50% workers adopt this route for entry into jobs[2]. According to OECD (1993, table 4.7), 23.6% of the young workers in France, 71.5% of those in Germany, and 67.1% of new hires in Japan receive formal training. Also, most of the European governments support life long learning by intervening through tax subsidies to learner and providers of training, grants, loans and training voucher programmes (OECD, 2004, 2005).

[1]See on https://www.td.org/research-reports/2018-state-of-the-industry. Similar results hold from 2018 Training Industry report: www.trainingmag.com.

[2]See, https://www.apprenticeship-toolbox.eu

Chapter 2

Training Investment and Rents in a Model with Endogenous Quits

2.1 Introduction

The human capital theory of Becker (1962) predicts that in competitive markets, firms cannot invest in worker's training that is completely transferable[1]. Studies have tested this prediction against different data sets and the conclusion is that seemingly general trainings are partly financed by employers (Loewenstein and Spletzer, 1998, Parent, 1999, Pischke, 2001, Booth and Bryan, 2002). Firms are found to invest in general training which takes place without any explicit long-term contract and are verifiable in the market[2]. To reconcile these findings, studies have introduced different extensions to the Becker's theory such as asymmetric information among firms, job-specific skills along with general skills, moral hazards, unions and minimum wages (Acemoglu and Pischke, 1998, 1999, Kessler and Lülfesmann, 2006,

[1] There are two types of human capitals in the context of training. Firm-specific training provides a worker with firm-specific skills, that is, skills that will increase her productivity only with the current employer. On the other hand, general training will add to the worker's general human capital increasing her productivity with a range of employers. Becker (1962) shows that firm cannot finance general training in a frictionless market, as it cannot recover the costs of general training once the workers are able to move and their skills are portable.

[2] For a brief review of studies and specific examples of firm-financed general training without contract, see Acemoglu and Pischke (1998) and Kessler and Lülfesmann (2006).

Lazear, 2009). These extensions allow the firm to charge rent on general training in the post-training period, thus enabling the firm to recover the training costs.

This study highlights a key prediction from these explanations and tests it against empirical evidence. Then, it presents an extended model that explains the empirical evidence relatively better. In particular, it shows that irrespective of the non-competitive assumptions they have made, all the models that rely on bargaining process for wage setting and/or assume exogenous quit decision predict that the firm must charge rent not only on training but also on any other input which is technological complement of skills in the production function[3]. The empirical evidence shows that firms are consistently charging rents on training and capital. However, there seems to be no rent on working hours. In most of the estimations, the marginal increase in wage from an hour of work is at least as much as the marginal increase in value added from an additional hour of work. This finding is interesting and is important in much broader sense. Bargaining process such as Nash bargaining is widely used in labour economics as a tool of wage determination in an imperfectly competitive environment. But this, along with the complementarity assumption, cannot withstand the empirical finding that firm does not charge rent on work hours[4].

It is natural to expect that soft workers skills complement working hours as the only way to realize such skills is having increasing production per hour after getting such skills. If this is the case, then how can the evidence that firms charge rent on all inputs other than working hours be explained? This study shows that when the firm sets profit maximizing wage[5] and workers' quit decisions are endogenous, then the firm charges no rent on working hours even if work hours are complementary with trainings. The basic mechanism that forces the firm to pay back marginal value

[3]It is standard in macroeconomics, labour, trade and industrial economics to find theories based on complementarity of inputs in the production function.

[4]However, it is desirable to carry careful confirmation of such empirical findings from more detailed and high quality matched data sets as such finding has implications for many fields. For more details, see Section 2.3 below.

[5]Note that by profit maximising wage the study means wage setting where the worker has no bargaining power.

of work is the fear that charging rent will increase worker's incentive to leave, as the decision of labour supply and mobility lies with the worker. In such a setting, the firm charges rent on only those factors which the firm controls such as trainings and capital. Furthermore, it shows that a model where firm invests both in general and specific skills and quit decisions are endogenous can explain the empirical evidence relatively better than the asymmetric information explanations.

This channel can explain the data relatively better but is different from earlier explanations that rely on simultaneous existence of job-specific and general skills. Acemoglu and Pischke (1999) show that one reason firm can finance general training is if specific and general trainings are technological complements in the production function[6]. As discussed in next section, besides their assumption of complementarity between general and specific trainings, their model setup implies that the firm must charge rent on work hours as well along with rent on training. Kessler and Lülfesmann (2006) relax the assumption of complementarity of specific and general trainings to show that firm can finance general training if the wage setting process is based on outside option rather than on Nash bargaining. But the implication of their findings is exactly the same as in Acemoglu and Pischke (1999). The current study's results hold both under the assumption of technological complementarity between specific and general trainings as in Acemoglu and Pischke (1999) and without this assumption as in Kessler and Lülfesmann (2006). But unlike Kessler and Lülfesmann (2006), it does not impose any assumption on the wage setting process. In such a setting, the firm finances general training and worker is getting full marginal value of an additional work hour even if working hours are technological complements with other factors of production[7].

The rest of the study is organized as follow. Section 2.2 highlights existing models and their predictions. Section 2.3 carries empirical check of the predictions. Section

[6]Note that this paper lays-out the condition under which firm can finance general training and is an important contribution in training literature. The aforementioned mechanism is just one example from the paper.

[7]Furthermore, unlike Katz and Ziderman (1990), Acemoglu and Pischke (1998), and Lazear (2009), this model does not rely on any type of asymmetric information between incumbent firm and the potential employers.

2.4 contains setup of an extended model and its results. Section 2.5 further extends the model and compares the market economy with the social planner's problem. Section 2.6 concludes the study.

2.2 Existing Models of General Training Investment

A Number of explanations are offered to support the evidence of firm-financed general training. The basic condition and most of the institutional mechanisms that are enough to ensure the condition are explained in Acemoglu and Pischke (1999). This study reports these explanations and its empirical implications in this section and then test it against empirical evidence in the next section. Before going to specific institutional mechanisms, it first explains the condition which ensures investment in general training.

For simplicity, the study follows the existing literature and explains the condition with a two periods model. Productivity and human capital of the worker is assumed zero in period 1 for simplicity. In period 1, the worker gets wage W_1 and the firm decides investment in general training denoted by T_g. In period 2, the worker can either stay and get wage $W_2(T_g)$ or leave and get outside wage of $W_E(T_g)$. All agents are assumed risk neutral and there is no discounting. Each worker produces with the production function $f(T_g)$ which is increasing, continuous and concave in T_g, irrespective of the firm where she works since training is general and equally productive everywhere. The cost of training $C_g(T_g)$ is strictly increasing, differentiable and convex. Moreover, $f(0) = C_g(0) = 0$ and $C'_g(0) = 0$, where the zero first derivative of training cost ensures that a positive general training level is socially optimal. As in Acemoglu and Pischke (1999), from this period 2 Nash bargaining wage can be written as;

$$W_2(T_g) = W_E(T_g) + \beta[f(T_g) - W_E(T_g)], \tag{2.1}$$

where $\beta \in [0, 1]$ is the bargaining power of the worker and the firm's outside option is zero. Period 2 wage does not depend on training cost as it is incurred in period 1 and is sunk in period 2. Using Equation (2.1), profit of the firm becomes

$$\Pi(T_g) = (1 - \beta)[f(T_g) - W_E(T_g)] - C_g(T_g) - W_1. \tag{2.2}$$

The firm chooses T_g to maximise profit which gives the following first-order condition

$$(1 - \beta)[f'(T_g) - W'_E(T_g)] - C'_g(T_g) = 0. \tag{2.3}$$

This implies that the necessary condition for the firm to invest in T_g is $f'(T_g) > W'_E(T_g)$ and $\beta < 1$. Putting this condition in the wage Equation (2.1) gives $f'(T_g) > W'_2(T_g)$. This means that the firm invests in training only if it can charge rent on the productivity of such training; the marginal wage increase from training must be less than the marginal output increase from such a training. But the firm can charge rent on training only if the external wage of the worker is compressed; that is $f'(T_g) > W'_E(T_g)$.

Now the study explains the institutional mechanisms from the existing literature which ensures that the above condition holds and its implications for wage changes. Besides other common predictions from these models[8], one testable prediction can be drawn about the slope of wage function in terms of working hours, firm's capital stock, etc. by using one commonly used assumption from the literature. More specifically, if the assumption that different inputs are technological complements in

[8]For example, all these explanations imply that the quit probability decreases in general training, unlike Becker's theory which predicts that general training does not affect quit probability in perfectly competitive markets. Studies have supported the prediction of these non-competitive models by showing that quit probability decreases in off-the-job training, which is likely to lead to general skills (Parent, 1999, Dietz and Zwick, 2016).

the production function holds[9], then most of the models this study reviews predict that if the firm charges rent on general training then it must also charge rent on other complementary inputs as well. The study focuses on working hours to show this result from the existing models. For this, it endogenises labour supply in the existing models in the following sections.

2.2.1 Asymmetric Information with Exogenous Quits

The first and most cited institutional mechanism in the literature which ensures the condition for training investment is asymmetric information about worker's skills (Katz and Ziderman, 1990, Acemoglu and Pischke, 1998, Lazear, 2009). These studies assume that the incumbent employer has more information about its workers skills at the start of period 2 than the other potential employers in the outside market. As a result, the outside wage offer will be based on some average measure of skills. On the other hand, the incumbent firm knows the exact skills, so it can offer a wage bit above the outside offer but below the true productivity of the worker. This rent opportunity gives the incumbent firm enough incentives to invest in training even if it is completely general. This is briefly explained using the setup of Acemoglu and Pischke (1999) but adding labour supply to it. Assume the following production function;

$$f(T_g, \eta, n) = \eta T_g^{\alpha} n^{1-\alpha}, \tag{2.4}$$

where η is worker ability, n denotes work hours and $\alpha \in (0, 1)$. The firm can train workers in period 1 and such training is common knowledge in period 2. On the other hand, the firm which hires the worker in period 1 is assumed to know η at the end of period 1, but the outsiders do not have such information in period 2 (asymmetric information). To simply further, assume that η can take only two

[9] In terms of the production function $f(T_g, T_s)$, the two variables T_g and T_s are technological complements if $\frac{\partial^2 f(T_g,T_s)}{\partial T_g \partial T_s} > 0$. The assumption that physical capital, labour hours, human capital or worker soft skills and technology are technological complements is common and is widely used in macroeconomics, microeconomics, trade, industry and labour economics.

values; $\eta = 0$ with probability p and $\eta = 1$ with probability $1 - p$. Period 2 wage can be contingent on ability as the incumbent firm knows ability, $W_2(T_g, \eta, n)$, but outside offer only depends on training and labour supply. In such a setting, and assume no bargaining power of workers for the sake of understanding the role of exogenous quit, the incumbent firm offers $W_2(T_g, \eta = 0, n) = 0$ to low ability workers, and to the high ability workers it offers;

$$W_2(T_g, \eta = 1, n) = W_E(T_g, n), \tag{2.5}$$

taking n as given. Given the wage offer, the worker decides optimal labour supply in the second period by maximizing her net utility[10]. Also, suppose that the worker quits for exogenous reason with probability $q \in (0, 1)$. Thus, the outside market consists of workers who quit due to exogenous reasons and one who are laid-off. This means that $W_E(T_g, n)$ is positive since some high ability workers leave with probability q. All low ability workers must leave since they get zero if stay. The outside wage must be equal to the expected productivity of workers found in the market as given below

$$W_E(T_g, n) = \frac{q(1-p)T_g^{\alpha} n^{1-\alpha}}{p + q(1-p)}. \tag{2.6}$$

The profit function and first order condition for training become

$$\Pi(T_g, n) = (1-q)(1-p)[f(T_g, \eta = 1, n) - W_E(T_g, n)] - C_g(T_g) - W_1, \tag{2.7}$$

$$\Pi'(T_g, n) = (1-q)(1-p)[f'(T_g, \eta = 1, n) - W_E'(T_g, n)] - C_g'(T_g) = 0. \tag{2.8}$$

[10] Worker chooses n to maximise $W_2(., ., n) - C(n)$, where $C(n)$ is utility costs of work. For detail sequence of events in such a setting, see Section 2.4.

From Equations (2.6) and (2.8), one can see that the condition for the firm investment in training $f'(T_g, n) > W'_E(T_g, n)$ is satisfied as $f'(T_g, n) = \alpha T_g^{\alpha-1} n^{1-\alpha}$ and $W'_E(T_g, n) = \frac{q(1-p)\alpha T_g^{\alpha-1} n^{1-\alpha}}{p+q(1-p)}$. Investment in training is ensured by wage compression in the external market which happens due to asymmetric information about worker's ability.

Implication: The above wage function implies that the firm is also charging rent on working hours. To see this note that $f'(T_g, n) = (1-\alpha)T_g^{\alpha} n^{-\alpha} > \frac{q(1-p)(1-\alpha)T_g^{\alpha} n^{-\alpha}}{p+q(1-p)} = W'_E(T_g, n) = W'_2(T_g, n)$ since $\frac{q(1-p)}{p+q(1-p)} < 1$. This proves the claim that the firm charges rent on working hours as well. However, this result critically depends on the complementarity assumption in the production function. If the production function is instead $f(T_g, \eta, n) = \eta + T_g + n$. Then the outside wage offer becomes $W_E(T_g, n) = \frac{p(T_g+n)+q(1-p)(1+T_g+n)}{p+q(1-p)}$. From this one can see that the slope of wage function must be equal to the slope of the production function in terms of both working hours and training.

2.2.2 Firm-Specific and General Trainings with Wage Bargaining

One reason presented in the literature as to why firms invest in general training is simultaneous existence of general and specific skills (Acemoglu and Pischke, 1999, Kessler and Lülfesmann, 2006). Suppose that the production function is $f(T_g, T_s, n)$, where T_s denotes level of firm-specific training. Since T_s is effective only in the incumbent firm, the outside productivity of a worker leaving in period 2 becomes $f(T_g, 0, n)$. With perfect competition in the external market, the outside wage becomes $W_E(T_g, n) = f(T_g, 0, n)$. The period 2 Nash bargaining wage and the profit functions become

$$W_2(T_g, T_s, n) = \beta f(T_g, T_s, n) + (1-\beta) f(T_g, 0, n), \tag{2.9}$$

$$\Pi(T_g, T_s, n) = (1-q)(1-\beta)[f(T_g, T_s, n) - f(T_g, 0, n)] - C_g(T_g) - W_1. \qquad (2.10)$$

From this, one can see that the condition for general training investment $f'(T_g, T_s, n) > W'_E(T_g, n)$ holds only when T_g and T_s are complements in the production function. Kessler and Lülfesmann (2006) relax this complementarity assumption and show the above result by replacing Nash bargaining with outside option criterian for wage setting. In Kessler and Lülfesmann (2006), the production function becomes $f(T_g, T_s, n) = g(T_g, n) + s(T_s, n)$ and wage is set as $W_2(T_g, T_s, n) = \beta f(T_g, T_s, n) = \beta[g(T_g, n) + s(T_s, n)]$.

Implication: Again, complementarity of n and T_s implies that the derivative of (2.9) with respect to n must satisfy $f'(T_g, T_s, n) > W'_2(T_g, T_s, n)$. But this result depends on complementarity assumption, otherwise the derivatives of the two functions are equal. In the case of Kessler and Lülfesmann (2006), the result holds even without complementarity assumption between n and T_s.

Remarks: Note that in wage bargaining, the parties may take into account the possibility that the worker's optimal work hours can be different from n if she leaves. But this will not change the above result. To see this, let n_e denotes working hours in the external market and n_i denotes working hours with the incumbent firm. Then if the worker leaves she will maximise $f(T_g, 0, n_e) - C(n_e)$. This will determine her equilibrium labour supply n_e^* in the external market. This will be taken as fixed in the bargaining and the bargaining wage now becomes $W_2(T_g, T_s, n_i) = \beta f(T_g, T_s, n_i) + (1-\beta) f(T_g, 0, n_e^*)$. If the worker stays, she chooses n_i to maximise this wage. The second term in this wage function is a fixed constant in the second period, which the worker gets as a partial reward for her outside worth. Now, the claim that the firm charges rent on working hours is straightforward as $f'(T_g, T_s, n_i) > W'_2(T_g, T_s, n_i) = \beta f'(T_g, T_s, n_i)$ for $\beta < 1$.

2.2.3 Asymmetric Information with Endogenous Quits and no Bargaining

This setup is based on Acemoglu and Pischke (1998) but with endogenous labour supply. Suppose the setup is the same as in Section 2.2.1. In addition, assume that the worker can get a utility shock θ with known distribution function $D(\theta)$. Then, it is optimal for the worker to quit if $W_E - C(n_e) + \theta - W_2 + C(n_i) \geq 0$, where $C(n_e)$ shows utility costs of work if the worker works in the external market and $C(n_i)$ shows similar costs if the worker stays in the incumbent firm. This gives a quit rate $q = 1 - D[W_2 - C(n_i) - W_E + C(n_e)]$. If wage is not contingent on ability, then the wage function must look something like (Acemoglu and Pischke, 1998, eq. 3)

$$W_2(T_g, n) = \tilde{\eta} T_g^{\alpha} n^{1-\alpha}, \tag{2.11}$$

where $\tilde{\eta}$ is the minimum cut-off ability for lay-off. However, if wage is contingent on ability as well then the result changes. To show this, assume that the distribution function $D(.)$ is uniform over $[0, 1]$. Then period 2 profit from the high ability worker becomes

$$\Pi_2(\eta = 1) = (1 - p)[W_2 - C(n_i) - W_E + C(n_e)][f(T_g, \eta = 1, n_i) - W_2]. \tag{2.12}$$

Maximizing this with respect to W_2 gives

$$W_2(T_g, \eta = 1, n_i, n_e) = \frac{T_g^{\alpha} n_i^{1-\alpha} + W_E - C(n_e) + C(n_i)}{2}. \tag{2.13}$$

The outside wage is determined in a similar fashion as in Section 2.2.1. The worker will maximise this against the utility costs of work if she decides to stay.

Implication: If the wage is not contingent on ability as in (2.11) then the firm is charging rent on working hours as well along with rent on training. But when wage is determined as in (2.13) then $W_2'(T_g, \eta = 1, n_i, n_e) = f'(T_g, \eta = 1, n_i)$ in terms of labour supply at the optimum. This can be shown by using first-order conditions for labour supply in the incumbent firm and in the external market. For example, the first-order condition for labour supply in the incumbent firm gives $(1-\alpha)T_g^{\alpha}n_i^{-\alpha} = C'(n_i)$. Thus, such a specification can explain the finding that the firm does not charge rent on working hours even if it complements training. However, for this asymmetric information explanation to be plausible, there should be a close and improving association in the productivity and wages of workers overtime. The reason is that the firm here rewards the workers for the skills which are even not visible to the external market as soon as the firm gets knowledge of such skills. Section 2.3 presents some evidence that shows that the association between wages and productivity remains the same overtime, which goes against the asymmetric information explanation.

2.2.4 Other Institutional Mechanisms

The condition for firm's investment in general training can also hold due to efficiency wages, unions and minimum wage laws as well, as is shown in Acemoglu and Pischke (1999). This study considers the case of efficiency wage here. Suppose the worker can produce $f(T_g)$ by putting efforts which cost e to the worker. If the worker does not put such efforts, her productivity remains zero in period 2. Suppose that the probability of being caught when not putting effort is p. Suppose also that the exogenous quit probability is zero, $q = 0$. Under this setting, the incentive compatibility for putting efforts is $W_2 - e \geq (1-p)W_2$. As in Acemoglu and Pischke (1999), if the worker quit she incurs fixed cost of $\triangle > 0$. This implies a participation constraint of $W_2 \geq f(T_g) - \triangle$. Then the optimal wage offer becomes

$$W_2(T_g) = max\left[f(T_g) - \triangle, \frac{e}{p}\right]. \tag{2.14}$$

In this setting, if wage is $W_2(T_g) = f(T_g) - \triangle$, then the slope of the wage function is equal to the slope of the production function with respect to all the variables. However, this is not supported in any of the empirical evidence presented below. On the other hand, if wage is $W_2 = \frac{e}{p}$ or $W_2(T_g) = \frac{e(T_g)}{p}$, then investment in training is possible only if wage is less steeply sloped in training than $f(T_g)$ is. But this leads to same results about wage slope in working hours as in the cases of asymmetric information and firm-specific skills. The results are the same in the cases of minimum wage and union bargaining. So, the testable empirical prediction from the above models is as summarised in the following proposition:

Proposition 1. *Given that;*

(i) Work hours, n, are technological complements of other inputs,

(ii) Wage is set through bargaining and/or worker's quit decision is exogenous,

(ii) Or wage is not contingent on the input which is a source of asymmetric information, then

the existing models which ensure $\frac{\partial f(T_g,.,.,n)}{\partial T_g} > \frac{\partial W_2(T_g,.,.,n)}{\partial T_g}$ *also imply* $\frac{\partial f(T_g,.,.,n)}{\partial n} > \frac{\partial W_2(T_g,.,.,n)}{\partial n}$.

In the next section, empirical evidence are presented which suggest that firms consistently charge rent on training, thus confirming the first part of Proposition 1. However, no evidence can be provided about firms charging rent on working hours, negating the second part of Proposition 1. Thus, all the above wage functions, except the one in (2.13) lack clear empirical support. This study puts further evidence which is not quite in support of wage function in (2.13) as well. Then it develops a model in Section 2.4 which can explain the evidence relatively better.

2.3 Existing and New Empirical Evidence

Fortunately, the increasing availability of matched employer-employee data sets can be used to assess these predictions. Few existing studies compared the coefficient on the training variable in the wage and production function regressions using such data sets. They unambiguously found that the firm charges rent on training; see Table A.1 in the appendix. Though they solely focused on training coefficient, one can compare the coefficients on working hours and capital stock as well from these studies[11]. Before going to discuss their results, the production function and wage equation they estimated are briefly highlighted. They considered a Cobb Douglas production function $Y_{it} = \hat{L}_{it}^{\gamma_l} K_{it}^{\gamma_k} e^{\xi_{it}} e^{\epsilon_{it}}$, where Y_{it} represents firm's value added, $\hat{L}_{it}$ is the firm effective labour input, K_{it} is total capital stock and ξ_{it} represents firm's technical efficiency. If training is the only observable source of skill difference, then the effective labour can be written as $ln\hat{L}_{it} = lnL_{it} + \gamma_T \bar{T}_{it} + Z_{it}$ following Konings and Vanormelingen (2015), where $\bar{T}_{it}$ represents average training intensity of firm's i workforce and Z_{it} is unobserved labour force ability of the firm. Thus the logarithmic production function becomes

$$lnY_{it} = \gamma_0 + \gamma_k lnK_{it} + \gamma_l lnL_{it} + \gamma_l \gamma_T \bar{T}_{it} + \omega_{it} + \epsilon_{it}, \tag{2.15}$$

where ω_{it} represents a combination of technical efficiency and unobserved labour force quality. The corresponding descriptive or reduced form wage regression is written as;

$$lnW_{it} = \psi_0 + \psi_k lnK_{it} + \psi_l lnL_{it} + \psi_T \bar{T}_{it} + \psi_z Z_{it} + \nu_{it}, \tag{2.16}$$

where W_{it} represents the firm's wage bill. Existing studies have estimated these two equations mainly with Generalised Method of Moments (GMM), and their results are highlighted in Table A.1 in the appendix[12]. Two of the papers, Conti (2005) and Colombo and Stanca (2008), use firm level data for analysis whereas Dearden et al. (2006) use industry level data. Training intensity of a firm is measured as a ratio of the number (hours) of trained workers to the total number of workers (hours) in these papers[13].

[11]Exception is Konings and Vanormelingen (2015) as they include capital-labour ratio only in the wage regression.

[12]Konings and Vanormelingen (2015) results about the coefficient of training are the same. The current study does not report their results as they are not including work hours variable separately in the wage equation. Using their data, this study reproduces their results later by including the relevant variables.

[13]They include other controls like age, education etc. but this study reports the coefficients of interest only.

From Dearden et al.'s (2006) results in Table A.1, one can see that the coefficients on training, capital and R&D are low in the wage regression as compared to production regression both with random-effects, system GMM without common factor restriction and system GMM with common factor restriction[14]. This is consistent so far with the prediction of the above non-competitive models when one keeps the assumption of complementarity of inputs in production. However, the coefficient of working hours is higher in the wage regression compared to the production function regression in two out of three estimations. Moreover, in the case where the coefficient is higher in the production function, the difference is not as big as it is in the other two columns. From Conti (2005) and Colombo and Stanca (2008), again the coefficients on training, capital and R&D are low in the wage regression as compared to production regression in all estimations[15]. However, the coefficient on working hours is higher in the wage regression in some of the estimations and lower in others, when compared to working hours coefficient in the production function. Although the coefficient of work hours is at least as high in the wage regression as in the production regression in most of the estimations, the results are not unambiguous at best. To get a better picture, the current study further investigates this in the next subsection.

2.3.1 Data and Estimation Strategy

To further check whether firms are charging rent on working hours, this study carries out more estimations. For this purpose, it uses the data and estimation strategy adopted by Konings and Vanormelingen (2015). The data they use is from the Belfirst database and includes the income statements of all Belgian's incorporated firms. This data covers 1997-2006 period and has information of firm's value added,

[14] This study does not report fixed-effects results from Dearden et al. (2006) because it is the same as system GMM results. Moreover, it reports results where the same instruments are used in the production and wage regressions, and where all specification tests are satisfied.

[15] Note that FE stands for fixed-effects and $F-GMM$ stands for first difference GMM. Moreover, this study is not reporting the current coefficient of R&D from Conti (2005) as many think that R&D investment can affect output with certain lags only (Dearden et al., 2006). Moreover, it does not report growth estimations and one where the specification tests are not satisfied.

number (and hours) of full time equivalent workers, material costs, labour costs, capital stock and workers' formal training. Each firm provides information on both the number of trained workers and hours of training along with cost of training for each year. The data has information on entry and exit of new workers, their qualifications and positions as well. However, this and material costs information are available only for the large firms[16].

Regarding estimation strategy, the input choices of firms are likely to be correlated with the unobserved productivity ω_{it} in Equation (2.15). Unlike the studies reported in Table A.1, Konings and Vanormelingen (2015) follow the procedure suggested by Ackerberg et al. (2015) to control for such endogeneity of inputs. Ackerberg et al. (2015) is based on the idea that inputs decisions carry information about ω_{it}. More specifically, suppose that the material costs are function of unobserved productivity and other inputs as $lnM_{it} = f(\omega_{it}, lnL_{it}, lnK_{it}, \bar{T}_{it})$. If this function is monotonically increasing in ω_{it}, conditional on lnL_{it}, lnK_{it} and $\bar{T}_{it}$, then it can be inverted and ω_{it} can be replaced with observables in the production equation as;

$$lnY_{it} = \gamma_k lnK_{it} + \gamma_l lnL_{it} + \gamma_l \gamma_T \bar{T}_{it} + f^{-1}(lnM_{it}, lnL_{it}, lnK_{it}, \bar{T}_{it}) + \epsilon_{it}. \quad (2.17)$$

Equation (2.17) is estimated with polynomials of order four in labour, materials, training and capital to proxy the inverse material demand. From this, one eliminates the part of value added that results due to unanticipated shocks at time t and measurement errors as $\hat{\varphi}_{it} = \gamma_k lnK_{it} + \gamma_l lnL_{it} + \gamma_l \gamma_T \bar{T}_{it} + f^{-1}(lnM_{it}, lnL_{it}, lnK_{it}, \bar{T}_{it})$. Now, to identify the coefficients γ_k, γ_l and $\gamma_l \gamma_T$, one needs three independent moment conditions in the second stage. Assume that ω_{it} follows first-order Markov process as $\omega_{it} = g_t(\omega_{it-1}) + \zeta_{it}$, where ζ_{it} is mean independent of all information known at $t-1$. It is standard to expect that capital is not correlated with ζ_{it} since capital is determined in advance. Konings and Vanormelingen (2015) report that training

[16]For more details on data and descriptive statistics, see Table 1 and Appendix B of Konings and Vanormelingen (2015).

decisions take place one year in advance as per the information from many managers in Belgium. Konings and Vanormelingen (2015) also argue that, due to high costs of hiring and firing, employment level decisions are not spontaneous[17]. With these timing assumptions, one can get the following three independent moment conditions

$$E[\zeta_{it}\begin{pmatrix} lnL_{it} \\ lnK_{it} \\ \bar{T}_{it} \end{pmatrix}] = 0. \tag{2.18}$$

To recover the implied ζ_{it}, for any candidate values of $(\gamma_k, \gamma_l, \gamma_l\gamma_T)$, one needs to first estimate $\omega_{it}(\gamma_k, \gamma_l, \gamma_l\gamma_T) = \hat{\varphi}_{it} - \gamma_k lnK_{it} - \gamma_l lnL_{it} - \gamma_l\gamma_T\bar{T}_{it}$. Next, non-parametrically regressing $\omega_{it}(\gamma_k, \gamma_l, \gamma_l\gamma_T)$ on its lag value and constant gives $\zeta_{it}(\gamma_k, \gamma_l, \gamma_l\gamma_T)$. Once this is estimated, one formulates sample analogue of (2.18) and minimises it to get γ_k, γ_l and $\gamma_l\gamma_T$. After estimating the production function, the ω_{it} is used as a regressor in the wage equation. The idea is to control for unobserved labour quality so as to consistently estimate the wage regression[18].

2.3.2 Results

The results are shown in Tables 2.1 and 2.2. Unlike earlier studies, this study carries the estimations with total quantities, such as firm's total value added and total wage bill instead of per worker terms[19]. The current study uses Konings and Vanormelingen (2015) definition of training throughout. Training is measured as a ratio of trained workers to the total number of workers in a firm in both tables except Columns 2-3 of Table 2.2. In these columns, training is defined as a ratio of training

[17] However, to make sure that this assumption is plausible, estimation is carried with lag labour as well.

[18] If ω_{it} is different from Z_{it} only due to industry and time effects, then including time and industry dummies along with ω_{it} can consistently estimate the wage regression. But other sources of significant difference between ω_{it} and Z_{it} can be a source of bias in the wage regression.

[19] The results are essentially the same whether one uses total or per worker quantities. In per worker estimations, the coefficients on inputs can turn out negative in many cases. This is not surprising as it means that the average productivity of, say, workers (hours) is falling with increasing employment. However, this may surprise many at first glance and makes the comparison between the coefficients in the production versus wage equations difficult. In estimations with total quantities, this must not happen.

Table 2.1: Base Level Estimations

	Overall			Manufacturing		Non-Manufacturing	
	OLS	OLS	ACF	OLS	ACF	OLS	ACF
Production function							
Employment	0.785*	0.747*	0.764*	0.767*	0.792*	0.735*	0.751*
	(0.001)	(0.004)	(0.008)	(0.008)	(0.016)	(0.005)	(0.010)
Capital	0.165*	0.123*	0.088*	0.151*	0.129*	0.115*	0.081*
	(0.001)	(0.003)	(0.004)	(0.006)	(0.008)	(0.003)	(0.004)
Training	0.460*	0.315*	0.243*	0.300*	0.215*	0.301*	0.258*
	(0.008)	(0.010)	(0.010)	(0.014)	(0.016)	(0.012)	(0.013)
γ_T	0.586*	0.422*	0.318*	0.391*	0.272*	0.410*	0.343*
	(0.010)	(0.014)	(0.014)	(0.020)	(0.021)	(0.017)	(0.018)
Wage equation							
Employment	1.006*	0.970*	0.976*	0.974*	0.982*	0.966*	0.971*
	(0.001)	(0.003)	(0.004)	(0.004)	(0.006)	(0.004)	(0.005)
Capital	0.029*	-0.004*	-0.017*	0.025*	0.017*	-0.012*	-0.024*
	(0.001)	(0.002)	(0.002)	(0.003)	(0.004)	(0.002)	(0.002)
Training	0.352*	0.250*	0.225*	0.215*	0.185*	0.252*	0.236*
	(0.006)	(0.007)	(0.007)	(0.010)	(0.009)	(0.008)	(0.008)
Productivity	-	-	0.356*	-	0.352*	-	0.352*
			(0.006)		(0.009)		(0.007)
$\psi_T = \gamma_T$							
χ^2	958.8	257.0	69.1	128.1	26.7	157.3	59.6
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$\psi_l = \gamma_l$							
χ^2	46826.8	3849.6	1526.5	1053.1	308.3	2632.6	1135.2
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Number of obs	806035	73816	73816	23318	23318	50498	50498
Firms	136842	13746	13746	3878	3878	9868	9868

Note: Significance at 5% level is indicated by *. Standard errors are estimated with 500 bootstrap iterations. The training coefficient in the production function gives $\gamma_l \gamma_T$. Dividing this by labour coefficient gives γ_T. Finally, all estimations include industry and time dummies.

hours to the total number of work hours in a firm. Labour is measured as the total number of full time equivalent employees in the firm except Columns 2-3 of Table 2.2, where labour is defined as total number of work hours. The dependent variable is firm's total value added in the production function and total wage bill in the wage equation in all estimations[20]. To make sure that ζ_{it} and lnL_{it} are not correlated,

[20]For manufacturing sectors, the variables are deflated using price deflator at the 4 digit NACE level from the European Statistical Office, whereas for the non-manufacturing sectors NACE 2 digit price deflator from the EU Klems database is used.

Columns 4-5 in Table 2.2 report estimation with lag employment level. Columns 6-7 of Table 2.2 introduce more variables. Following Konings and Vanormelingen (2015), schooling is defined as the share of entrants with high school or university degrees out of the total entry into a firm over the sample period. Similarly, contract type is defined as share of management plus employees staff out of the total workforce in each year. Given the definition of schooling, it also works as a firm level dummy variable. Finally, the tables report results for OLS for the overall sample, OLS estimation of the reduced sample of those firms which report material costs and estimations with Ackerberg et al. (2015). Moreover, Table 2.1 carries results for overall sample and for manufacturing and non-manufacturing sectors separately.

As indicated in Tables 2.1 and 2.2, all the coefficients take expected signs in both the wage and production regressions[21]. These estimations once again confirm that the firm is charging rent on training and capital. No matter which definition of training and labour one uses, and irrespective of the estimation method, the coefficients of training and capital are significantly high in the production function compared to that in the wage equation. However, the story is different for the coefficient of labour. Whether one uses total employment or the number of hours worked, the coefficient of labour is consistently high in the wage equation when compared to that in the production equation. The difference is significant for all specifications when one uses total employment as a measure of labour. However, when one uses work hours as a measure of labour, the difference is not significant statistically in the ACF estimations. Thus, these results show that the marginal increase in wage due to one more hour of work is at least as much as the marginal increase in value added due to that hour of work, after controlling for other factors. If the true coefficient of working hours is really high in the wage equation or is equal to the one in the production function regression, along with the findings that the firm charges rent on capital, training and R&D, it has important implication for the

[21] The only exception is the sign of capital in some of the wage regressions. The sub-sector analysis in Table 2.1 shows that this is happening due to non-manufacturing sector. However, when I include additional variables in Table 2.2, the sign becomes positive.

above mentioned theoretical settings. One possible implication is that capital and R&D are technological complements with training but working hours are not. But this is very unlikely explanation as discussed before. If this is not the case, then the setting with asymmetric information, endogenous quits and where wage is contingent on unobserved ability (see Subsection 2.2.3) can explain the empirics relatively better than the other settings. This study checks this asymmetric information based explanation further in the following analysis.

If the incumbent firm is gradually collecting information about abilities or other skills of workers and incorporates it into wage setting, as highlighted in the asymmetric information story above and detailed in Acemoglu and Pischke (1998), then the association between the worker productivity and wage should increase overtime. This is because, at the time of hire, the firm can make decisions only on observables such as education. Overtime, when the firm gets information on the workers' true skills, those workers' wages should increase more who are more productive relative to less able workers. This means that the association in workers productivity and wages should be high in later years as compared to the entry year. The data used in Konings and Vanormelingen (2015) has information on the number of entrants with their qualifications in each year. Similarly, the data gives information on the number of quitters and their education. From this data, one can check how such entry and exit affects the firm total value added and wage bill. According to the asymmetric information story, the association in the productivity and wage changes should be high at the exit point as compared to the association at the time of entry[22].

These results are shown in Table A.2 in the appendix. Columns 2-5 with row university show how new entrant with a university degree affects value added and wage bill of a firm. The last four columns show the effects of worker's quit with a university degree on wage bill and the total value added. The two other rows show the wage and productivity effects of entry and exit of workers with high school and secondary schooling. The coefficients for the three education categories in the wage

[22]The exit does not include dismissals and retirements.

Table 2.2: Further Robustness Check

	Overall with Hours		Overall with lag Labour		Overall	
	OLS	ACF	OLS	ACF	OLS	ACF
Production function						
Labour	0.770*	0.586	0.751*	0.810*	0.770*	0.785*
	(0.004)	(0.416)	(0.004)	(0.016)	(0.004)	(0.023)
Capital	0.122*	0.113*	0.122*	0.070*	0.141*	0.116*
	(0.003)	(0.016)	(0.003)	(0.004)	(0.002)	(0.008)
Training	0.008*	0.006*	0.310*	0.223*	0.162*	0.131*
	(0.000)	(0.001)	(0.010)	(0.011)	(0.008)	(0.010)
γ_T	0.011*	0.010	0.413*	0.276*	0.211*	0.166*
	(0.000)	(0.009)	(0.014)	(0.015)	(0.011)	(0.014)
Schooling	-	-	-	-	0.385*	0.656
					(0.016)	(997.8)
Contract type	-	-	-	-	0.482*	0.330*
					(0.014)	(0.051)
Wage equation						
Labour	0.996*	0.933*	0.971*	0.992*	0.985*	0.989*
	(0.004)	(0.142)	(0.003)	(0.007)	(0.002)	(0.007)
Capital	-0.003	-0.006	-0.003	-0.022*	0.012*	0.005
	(0.002)	(0.006)	(0.002)	(0.002)	(0.001)	(0.003)
Training	0.006*	0.005*	0.246*	0.215*	0.118*	0.109*
	(0.000)	(0.000)	(0.007)	(0.007)	(0.005)	(0.005)
Productivity	-	0.341*	-	0.356*	-	0.281*
		(0.007)		(0.006)		(0.005)
Schooling	-	-	-	-	0.391*	0.467
					(0.011)	(279.2)
Contract type	-	-	-	-	0.384*	0.341*
					(0.009)	(0.017)
$\psi_T = \gamma_T$						
χ^2	294.3	0.3	214.9	24.9	91.4	22.1
p-value	0.000	0.592	0.000	0.000	0.000	0.000
$\psi_l = \gamma_L$						
χ^2	9193.9	1.6	3778.8	301.9	3749.0	151.3
p-value	0.000	0.207	0.000	0.000	0.000	0.000
Number of obs	72803	72803	63134	63134	72491	72491
Firms	13691	13691	12520	12520	13241	13241

Note: Significance at 5% level is indicated by *. Standard errors are estimated with 500 bootstrap iterations. The training coefficient in the production function gives $\gamma_l\gamma_T$. Dividing this by labour coefficient gives γ_T. Finally, all estimations include industry and time dummies.

and production regressions are used to estimate the association between the two for the entry and exits. From this, one can see that there is no major difference in the wage-productivity association for the entrants versus quitters. The wage-productivity association is at least as strong at the time of entry as it is at the time of exit. The advantage of this last exercise is that it is not likely to happen due to estimation bias. If any bias exists, it should affect the entry coefficients as much as it does the exit coefficients. This last finding seems to go against the explanation based on asymmetric information story. However, more detailed and careful empirical analysis with a rich data would be desirable, as this finding has implications for the entire field of labour economics. Instead of going into that dimension, in the next section, this study develops a model which can explain such finding relatively better.

2.4 Model Setup and Results

This model makes two contributions. First, it develops a setting where the firm is charging rent on training but has no incentives to charge rent on working hours. Second, it provides an alternative explanation for why firms invest in a completely general training. More specifically, it relaxes both Acemoglu and Pischke's (1999) assumption of technological complementarity between general and specific training and Kessler and Lülfesmann's (2006) assumption of wage setting, and shows that the firm still has incentives to finance general training given that the worker uses both general and specific skills simultaneously in a given task and/or efforts are endogenous.

2.4.1 The Basic Framework

Once again, the notations used and the basic model structure is briefly described to refresh memory. The study starts with a simple model by assuming that the world lasts for only two periods, all firms and workers are risk neutral and there is no discounting between periods (Acemoglu and Pischke, 1998, Kessler and Lülfesmann, 2006, Lazear, 2009). Output and worker's human capital is zero in period 1. During period 1, firms can offer both specific and general trainings to workers they hire and the levels of such trainings are denoted by T_s and T_g, respectively. Period 1 is allocated to trainings entirely and the worker works in period 2. The firm bears the direct costs of specific and general trainings denoted by $C_s(T_s)$ and $C_g(T_g)$, respectively. At the end of period 1, the training firm, external/outside market and worker know the level of general training. Worker can also get utility shock in the training firm[23] in period 1, denoted by θ with expected value of $E(\theta) = 0$. The training firm and the external market can only know the distribution function $D(\theta)$, and $D(\theta)$ is assumed to be continuously differentiable (Black and Loewenstein, 1997, Acemoglu and Pischke, 1998). This creates uncertainty for the firm, as in the case of a negative utility shock the worker can leave the firm in period 2.

In period 2, the value of a trained worker at the training firm is assumed to depend on general training, specific training and working hours n and is denoted by $f(T_g, T_s, n_i)$. The value of the trained worker in the external market becomes $f(T_g, 0, n_e)$ as specific training is worthy only at the training firm. Note that n_e represents working hours when the worker chooses to leave, and it can be different from n_i as the value of worker is different at the training firm versus in the external market. The total time M will be allocated as $M = l_i + n_i$ or $M = l_e + n_e$, depending on where the worker works. The external market is perfectly competitive so that the worker will get her full worth if she leaves the training firm. Work involves utility costs denoted by $C_i(n_i)$ and $C_e(n_e)$ for the training firm and in the external market, respectively. Thus, the external utility of the worker becomes $f(T_g, 0, n_e) - C_e(n_e)$. Given period 2 wage offer W_2 and external value of the worker, the worker decides where to work. Labour supply is determined by the worker utility maximization, given quit decision. Finally, one way is to continue with the production function $f(T_g, T_s, n)$ and assume that T_g and T_s are technological complements as in Acemoglu

[23]The worker may not go along with the boss, colleagues, can dislike the work environment or the city and vice versa.

and Pischke (1999). However, this study addresses Kessler and Lülfesmann's (2006) concerns also and relaxes the assumption of complementarity between T_g and T_s. Following Kessler and Lülfesmann (2006), this study assumes that the productivity of a skilled worker is additively separable in T_g and T_s as;

$$f(T_g, T_s, n_i) = G(T_g, n_i) + S(T_s, n_i), \tag{2.19}$$

where $G(.)$ and $S(.)$ are values of general and specific skills, respectively. Thus, if the worker decides to leave in period 2, her value will be $f(T_g, 0, n_e) = G(T_g, n_e)$. This means that T_g is fully transferable whereas T_s is not transferable at all. When one is not sure about whether T_g and T_s are technological complements, the idea of additive separability assumption is to check whether T_g and T_s can be incentive complementary for the firm or not[24]. In Kessler and Lülfesmann (2006), T_g and T_s are additively separable and the incentive complementarity in their study arises due to their assumed value sharing assumption. This study imposes no assumption about wage setting; rather wage is determined by profit maximization behavior of the firm. Incentive complementarity between T_g and T_s can arise in this model due to n term in the production function. But n is a decision variable of the worker rather than the firm, and also worker has to face $C(n)$ costs in exerting n.

Assumptions

1. The function $f(T_g, T_s, n)$ is twice continuously differentiable, strictly increasing and concave in T_g, T_s and n, respectively. Furthermore, it satisfies the Inada type conditions $\lim_{j\to 0} \frac{\partial f(T_g,T_s,n)}{\partial j} = \infty$ and $\lim_{j\to \bar{j}} \frac{\partial f(T_g,T_s,n)}{\partial j} = 0$ for $j \in (T_g, T_s, n)$, where $\bar{j}$ represents the upper bounds on trainings and work, respectively.

[24]Note that technological complementarity holds when two inputs are direct complements in production function, i.e. $\frac{\partial^2 f(T_g,T_s,n_i)}{\partial T_s \partial T_g} > 0$. Incentive complementarity means that it pays to the profit maximiser firm to increase the level of one input when it increases the level of the other even if the two inputs are additively separable in the production function, i.e. when $\frac{\partial^2 f(T_g,T_s,n_i)}{\partial T_s \partial T_g} = 0$ but cross partials of profit function is positive.

2. The cost functions $C_v(j)$ are strictly increasing and convex for all $j \in (T_g, T_s, n)$ and $v \in (g, s, i, e)$. Moreover, $\frac{\partial C_v(j=0)}{\partial j} = 0$. Also, $\frac{\partial C_i(n)}{\partial n}$, respectively $\frac{\partial C_e(n)}{\partial n}$, is finite for all $n \in (0, \bar{n})$.

3. Work is complementary in production with the general and specific trainings, respectively, i.e. $\frac{\partial^2 G(T_g,n)}{\partial n \partial T_g} > 0$ and $\frac{\partial^2 S(T_s,n)}{\partial n \partial T_s} > 0$.

4. There is perfect competition in both the periods and all parties observe general training.

5. The survivor function $[1 - D(\theta)]$ is strictly log concave, i.e. $\frac{\partial^2(\log[1-D(\theta)])}{\partial \theta^2} < 0$.

6. When working in the incumbent firm in period 2, the worker uses general and specific skills simultaneously in a given task.

The first assumption is standard and ensures well-behaved solution. The second part of Assumption 2 implies that it is socially optimal to have positive values of T_g, T_s and n, and the first derivative of utility costs is well-defined for all possible n values. Assumption 3 is most natural and plausible assumption, and is standard across most disciplines. Assumption 4 implies that the worker wage in the external market is $f(T_g, 0, n_e) = G(T_g, n_e)$. Assumption 5 ensures well-behaved second order conditions and comparative statics, and is common in such studies (Acemoglu and Pischke, 1998). This assumption is satisfied by almost all of the distributions which has both positive and negative values in the support; including distributions such as Uniform, Normal, Logistic, Laplace etc.

Assumption 6 means that general and specific skills are used simultaneously in a given task rather than using it in two completely different tasks. Thus, the worker only decides how many hours to work with the firm. Relaxing this assumption would mean writing (2.19) as $f(T_g, T_s, n_g, n_s) = G(T_g, n_g) + S(T_s, n_s)$. This means that the worker performs two different tasks in the incumbent firm, where one task needs only general skills and the other only specific skills. In that case, the worker can decide how much time to allocate for working with the general skills and how many hours to work with the specific skills task. In the real world, a clear distinction between general and specific skills may not be possible, and in most cases a worker uses many skills simultaneously while working. For example, if a worker is trained in general computer skills and on a specific software, which is used within the firm only, then the worker will use both the general computer and specific software skills simultaneously when working on the software. So, in such a situation the worker can decide on the total working hours only. Even in most multi-task jobs, the worker can decide total hours and the employer decides how the worker should allocate time between different tasks. This assumption only affects the firm decision to invest in general training. The result about the firm's rent charging behavior does not depend on this assumption in any way. In addition, in the extension of the model in Section 2.5, it is shown that all the results of this section can hold even when this assumption is relaxed.

Sequence of Events

1. At the beginning of period 1 firm hires workers and decides the level of trainings T_g and T_s. At the end of period 1 the training firm as well as the external market know the level of T_g.

2. A worker can get utility shock in the training/incumbent firm denoted by θ with expected value of $E(\theta) = 0$. All the firms only know the distribution function $D(\theta)$.

3. At the beginning of period 2 the incumbent firm offers wage W_2. External firms make wage offer W_E to those workers who are in the external market in period 2. There is perfect competition in the external market so that the worker will get $W_E = G(T_g, n_e)$. Worker bears utility cost $C_e(n_e)$, so that the expected worth of her external job is $G(T_g, n_e) - C_e(n_e)$.

4. Given the realization of θ and $G(T_g, n_e) - C_e(n_e)$, the worker will quit if period 2 wage W_2 and non-pecuniary benefits θ, net of $C_i(n_i)$, are less than her outside net value i.e.,

$$W_2 + \theta - C_i(n_i) < G(T_g, n_e) - C_e(n_e). \tag{2.20}$$

This equation thus guides worker's quit decision and will be used to determine worker's quit probability.

Equilibrium Concept

The equilibrium concept this study uses is an intuitive backward solution like in Black and Loewenstein (1997) or in Acemoglu and Pischke (1998). Being in period 1, the training employer will think about what level of W_2 should the firm offer to get maximum out of the worker. The best way for the firm is to offer W_2 which maximises the firm profit for a given labour supply (Acemoglu and Pischke, 1998). Once W_2 offer is there, the worker can decide on the labour supply from her utility maximization behavior. Knowing the best response labour supply of workers, for any given W_2 offer, the firm can decide on the training levels in period 1 by choosing trainings that give maximum profit. These steps are summarized as following

1. The firm offers W_2 which ensures maximum profit for the firm, for any given levels of T_g, T_s, n_i, and n_e.

2. Given the wage offer W_2, realization of θ and external offer W_E, the worker takes optimal quit decision on the basis of Equation (2.20). Given quit decision, the worker decides labour supply n_e or n_i by maximizing utility $W_E - C_e(n_e)$ or $W_2 - C_i(n_i)$. Note that the worker faces no uncertainty during her decisions.

3. Given this and the training costs, the firm will now choose T_g and T_s to maximise profit in period 1.

4. Period 1 wage is determined by free entry conditions.

2.4.2 Optimal Wage and Labour Supply

Note from above that when deciding on trainings, the firm will take into account its optimal wage offer. Second, as will get clear below, whether the firm takes into account the best response labour supply in step 3 or chooses trainings by taking labour supply as given will give same results. Also, note that the worker with a given T_g can choose different working hours depending on whether she stays or quits. The important point here is that the external wage is determined competitively. This implies that the external firm can earn zero profit by hiring the trained worker in period 2, so that the trained worker's external wage is $W_E = G(T_g, n_e)$. Thus, the worker quits if the realization of θ turns out to be less than the critical value $\{W_E - C_e(n_e) - W_2 + C_i(n_i)\}$, so the worker's quit probability is $D\left(W_E - C_e(n_e) + C_i(n_i) - W_2\right)$. Given this the employer's profit in period 2 is given by

$$\Pi_2 = [1 - D(W_E - C_e(n_e) + C_i(n_i) - W_2)]\{f(T_g, T_s, n_i) - W_2\}. \tag{2.21}$$

The wage offer that maximises Π_2 satisfies the first-order condition (Black and Loewenstein, 1997, Acemoglu and Pischke, 1998)

$$d(W_E - C_e(n_e) + C_i(n_i) - W_2)\{f(T_g, T_s, n_i) - W_2\} - [1 - D(W_E - C_e(n_e) + C_i(n_i) - W_2)] = 0,$$

where $d(W_E - C_e(n_e) + C_i(n_i) - W_2)$ is the density function. Solving this one can get second period optimal wage as

$$W_2^*(T_g, T_s, n_i, n_e) = f(T_g, T_s, n_i) - \frac{[1 - D(W_E - C_e(n_e) + C_i(n_i) - W_2^*)]}{d(W_E - C_e(n_e) + C_i(n_i) - W_2^*)}. \tag{2.22}$$

Note that the equilibrium wage is a function of both T_g, T_s and work along with quit probability[25]. Later on, one will see that labour supply also depends on trainings so that after putting optimal labour supply, wage will be a function of trainings only, i.e. $W_2^*(T_s, T_g)$, as in Acemoglu and Pischke (1998). In addition, wage is unambiguously less than the total output from the trained worker. Moreover, for the second-order condition, take the second derivative of (2.21) with respect to W_2

[25] In the inverse hazard function $\frac{[1-D(W_E-C_e(n_e)+C_i(n_i)-W_2^*)]}{d(W_E-C_e(n_e)+C_i(n_i)-W_2^*)}$ I use W_2^* instead of $W_2^*(T_g, T_s, n_i, n_e)$ just to save space.

and using the first-order condition one gets $-\frac{[1-D(.)]\acute{d}(.)}{d(.)} - 2d(.)$. Where $D(.)$ and $d(.)$ are short hands for $D(W_E - C_e(n_e) + C_i(n_i) - W_2^*)$ and $d(W_E - C_e(n_e) + C_i(n_i) - W_2^*)$, respectively, and $\acute{d}(.)$ is the first derivative of $d(.)$. Thus, the second-order conditions hold when $\frac{[1-D(.)]\acute{d}(.)}{[d(.)]^2} > -2$. Assumption 5 gives $\frac{[1-D(.)]\acute{d}(.)}{[d(.)]^2} > -1$, ensuring the second order conditions hold.

Now, when deciding on the level of T_g and T_s, the firm can proceed in two different ways. (1) The firm may ignore its current period's training decisions effects on period 2 labour supply. In this case, the firm treats n_i and n_e as given and chooses T_g and T_s to maximise profit. (2) The firm can take into account the effects of its today's training decisions on period 2 labour supply. In this case, the firm needs to take into account the best response labour supplies while deciding T_g and T_s. This study proceeds with the case 2 for the sake of better understanding[26]. To formulate the firm trainings choice problem, one needs to first get the worker best response to the firm wage offer. For this, the behavior of the optimal wage in (2.22) with respect to n_i is studied first.

Lemma 1. *Given the assumptions of the model, the optimal period 2 wage* $W_2^*(T_g, T_s, n_i, n_e)$ *is continuous in* n_i *and satisfies* $\lim_{n_i \to 0} \frac{\partial W_2^*(T_g,T_s,n_i,n_e)}{\partial n_i} = \infty$ *and* $\lim_{n_i \to \bar{n}_i} \frac{\partial W_2^*(T_g,T_s,n_i,n_e)}{\partial n_i} > 0$*. Furthermore,* $\lim_{n_i \to \bar{n}_i} \frac{\partial W_2^*(T_g,T_s,n_i,n_e)}{\partial n_i} < \lim_{n_i \to \bar{n}_i} \frac{\partial C_i(n_i)}{\partial n_i}$.

Proof. See the appendix. □

From Lemma 1, one can see that the slope of wage function starts at infinity and approaches $[1 + \frac{(1-D(.))\acute{d}(.)}{[d(.)]^2}]\frac{\partial C_i(n_i)}{\partial n_i}/[2 + \frac{(1-D(.))\acute{d}(.)}{[d(.)]^2}]$ as labour supply approach $\bar{n}_i$. Additionally, the slope must be negative at low level of labour supply. However, it can become positive at a high level of labour supply if the decreasing first term has been dominated by the increasing second term as is clear from Lemma 1. But, $\lim_{n_i \to \bar{n}_i} \frac{\partial W_2^*}{\partial n_i} = \lim_{n_i \to \bar{n}_i} \left[[1 + \frac{(1-D(.))\acute{d}(.)}{[d(.)]^2}]\frac{\partial C_i(n_i)}{\partial n_i}/[2 + \frac{(1-D(.))\acute{d}(.)}{[d(.)]^2}] \right] < \lim_{n_i \to \bar{n}_i} \frac{\partial C_i(n_i)}{\partial n_i}$. This implies that the marginal costs of work must lie above the marginal benefits of

[26] Both approaches give same results because of Envelope theorem. For more details, see appendix.

work in this range. Thus, there is only one crossing point where the marginal cost of work cut the slope of wage function from below.

The study now turns to optimal labour supply determination. Given the wage offer $W_2^*(T_g, T_s, n_i, n_e)$, the external wage offer W_E and the realization of θ, the worker is in a position to decide whether to stay or quit by simply following the rule in Equation (2.20). If the worker decides to stay, she solves the following problem to maximise utility

$$\underset{n_i}{max} U = W_2^*(T_g, T_s, n_i, n_e) - C_i(n_i). \tag{2.23}$$

Equating marginal cost and benefits of work, $\frac{\partial W_2^*(T_g,T_s,n_i,n_e)}{\partial n_i} = \frac{\partial C_i(n_i)}{\partial n_i}$, and cancelling common terms, one can get

$$\frac{\partial f(T_g, T_s, n_i^*)}{\partial n_i} = \frac{\partial C_i(n_i^*)}{\partial n_i}. \tag{2.24}$$

Given the assumptions of the model, the uniqueness of $n_i^*(T_g, T_s)$ is ensured. Similarly, if the worker decides to quit, her equilibrium labour supply will be $n_e^*(T_g)$ which is obtained by maximizing $\underset{n_e}{max} U = G(T_g, n_e) - C_e(n_e)$. Moreover, the non-negativity and time constraints on working hours are not binding due to assumptions of the model.

Proposition 2. *Given Assumption 3, utility cost functions $C_i(.) \equiv C_e(.)$ and that $T_s > 0$ holds, then the worker's equilibrium labour supply in the training firm must be strictly greater than her labour supply in the external market, i.e. $n_i^*(T_g, T_s) > n_e^*(T_g)$.*

Proof. This can be proved easily by contradiction. Suppose to the contrary $n_i^*(T_g, T_s) \leq n_e^*(T_g)$. Then by the concavity of $G(T_g, n)$ in n and first-order conditions for work, $\frac{\partial C_e(n_e^*(T_g))}{\partial n} = \frac{\partial G(T_g,n_e^*(T_g))}{\partial n} \leq \frac{\partial G(T_g,n_i^*(T_g,T_s))}{\partial n} < \frac{\partial C_i(n_i^*(T_g,T_s))}{\partial n}$. The last inequality holds due to the fact that $S(T_s, n)$ is strictly increasing in n, which in turn implies $\frac{\partial f(T_g,T_s,n)}{\partial n} = \frac{\partial G(T_g,n)}{\partial n} + \frac{\partial S(T_s,n)}{\partial n} > \frac{\partial G(T_g,n)}{\partial n}$ for all n. But given that $C_i(.) \equiv C_e(.)$ and are strictly convex, $\frac{\partial C_e(n_e^*(T_g))}{\partial n} < \frac{\partial C_i(n_i^*(T_g,T_s))}{\partial n}$ implies $n_i^*(T_g, T_s) > n_e^*(T_g)$. Hence a contradiction. $\square$

Proposition 2 holds due to the fact that at $T_s > 0$, marginal productivity from work is high in the training firm as compared to the external market, as work and specific training are complementary in production. If the worker decides to quit, her specific skills become worthless in the external market. This reduces her work hours productivity in the external market at the margin. Thus, it pays her to work less hours in the external market as a compared to her optimal work hours choice in the incumbent firm. The magnitude of the difference depends on the level of specific skills the worker get in period 1 and on the nature of work's utility cost functions. If the utility cost function is strongly convex, the difference may be small for example. Moreover, the equilibrium conditions for work also imply that in equilibrium the wage change in work hours must equal to the output change due to work[27]. Proposition 2 is an important finding and provides the basis for investment in general training by the employer even if T_g and T_s are not complements in the production. If the trained worker works more at the training firm, this provides an opportunity to the training firm to earn more rent on the trainings as training and work hours complement each other. This will get more clear in the following section.

2.4.3 Optimal Trainings Choices

Before analysing the wage effects of training and other factors, the study first turns to the trainings decision of the firm in period 1 to establish the basic point that the firm can invest in general training in such a setting. In the first period, the firm trains workers by spending $C_s(T_s)$ and $C_g(T_g)$, respectively. Additionally, it pays period 1 wage of W_1. As noted earlier, when deciding on trainings, the firm takes into account the best response labour supplies $n_i^*(T_g, T_s)$ and $n_e^*(T_g)$ and period 2

[27] This point is discussed later when the study analyses the slope of wage function.

optimal wage offer[28]. Given this, the overall profit and trainings decision problem can be set up as (Acemoglu and Pischke, 1998)

$$\max_{\{T_g,T_s\}} \Pi = [1-D(W_E-C_e(n_e^*)+C_i(n_i^*)-W_2^*)]\{f(T_g,T_s,n_i^*)-W_2^*\}-C_g(T_g)-C_s(T_s)-W_1. \tag{2.25}$$

Note that $W_E = G(T_g, n_e^*(T_g))$ and W_2^* is given by (2.22) after replacing n_i with $n_i^*(T_g, T_s)$. Thus, when choosing T_s and T_g, the firm now considers the direct effects of trainings on output and wage and the indirect effects that arise through labour supply. The trainings that maximise Π satisfy the first-order conditions

$$d(.)\left(\frac{\partial W_2^*}{\partial T_s} - \frac{\partial C_i(n_i^*)}{\partial T_s}\right)\{f(T_g,T_s,n_i^*)-W_2^*\}+\{\frac{\partial f(T_g,T_s,n_i^*)}{\partial T_s}-\frac{\partial W_2^*}{\partial T_s}\}[1-D(.)] = \frac{\partial C_s(T_s)}{\partial T_s}, \tag{2.26}$$

$$\begin{aligned} d(.)\left(\frac{\partial W_2^*}{\partial T_g} - \frac{\partial W_E}{\partial T_g} + \frac{\partial C_e(n_e^*)}{\partial T_g} - \frac{\partial C_i(n_i^*)}{\partial T_g}\right)\{f(T_g,T_s,n_i^*) - W_2^*\} = \\ \frac{\partial C_g(T_g)}{\partial T_g} - \{\frac{\partial f(T_g,T_s,n_i^*)}{\partial T_g} - \frac{\partial W_2^*}{\partial T_g}\}[1 - D(.)]. \end{aligned} \tag{2.27}$$

The understanding that W_2^* is a function of T_g and T_s, and W_E is a function of T_g is used in the derivations above. Also, utility cost of work changes in training due to the training effects on work, i.e. $\frac{\partial C(n_i^*(T_g,T_s))}{\partial T_g}$. The first-order conditions for specific training in (2.26) and general training in (2.27) show that the firm invests in trainings to the point where marginal costs of respective trainings are equal to the sum of respective net marginal benefits from trainings given quit probability, and the added effects of trainings on the probability of quit.

[28]Note that, depending on the trade-off between space limitations and better understanding, in rest of the study either $n_i^*(T_g, T_s)$ or the short hand n_i^* is used. Similarly, I will either use n_e^* and W_2^* or $n_e^*(T_g)$ and $W_2^*(T_g, T_s, n_i^*, n_e^*)$.

The effects of trainings arise directly due to the training terms in the production function and indirectly through the training effects on labour supply. For example, the effects of training change on output will be $\frac{\partial f(T_g,T_s,n_i^*)}{\partial T_g} = \frac{\partial G(T_g,n_i)}{\partial T_g}|_{n_i=n_i^*} + \frac{\partial f(T_g,T_s,n_i^*)}{\partial n_i}\frac{\partial n_i^*}{\partial T_g}$. Here, the first term shows the direct effects of training on worker's value for a given labour supply and the second term shows the effects which arise due to the training effects on working behavior. But using equilibrium conditions for work, i.e. $\frac{\partial W_2^*}{\partial n_i} = \frac{\partial C_i(n_i)}{\partial n_i}$ and $\frac{\partial W_E}{\partial n_e} = \frac{\partial C_e(n_e)}{\partial n_e}$, all these indirect effects cancel out[29]. The intuition behind all such cancellations is simple. Under the case of best response consideration, the firm realises the indirect effects of training on worker's value $\frac{\partial f(T_g,T_s,n_i^*)}{\partial n_i}\frac{\partial n_i^*}{\partial T_g}$. But then the firm is needed to also takes into account the wage change that is needed to induce the worker to put work $\frac{\partial W_2^*}{\partial n_i}\frac{\partial n_i^*}{\partial T_g}$. Given that the worker chooses labour supply at the point where $\frac{\partial W_2^*}{\partial n_i} = \frac{\partial C_i(n_i^*)}{\partial n_i}$, the marginal increase in wage due to work $\frac{\partial W_2^*}{\partial n_i}\frac{\partial n_i^*}{\partial T_g}$ must be equal to the marginal increase in the value of the worker due to work $\frac{\partial f(T_g,T_s,n_i^*)}{\partial n_i}\frac{\partial n_i^*}{\partial T_g}$. This can be seen from (2.23) and (2.24), which provides the basis for why firm is not charging rent on work, as will get clear later. Thus, these two forces cancel out and level of trainings has not been affected by the fact that whether the firm takes into account the worker best response labour supply or not when decides trainings. Thus, making use of (2.19) and (2.22) and equilibrium conditions for work, one can simplify (2.26) and (2.27) to get

$$[1 - D(.)]\frac{\partial S(T_s, n_i)}{\partial T_s}|_{n_i=n_i^*(T_g,T_s)} = \frac{\partial C_s(T_s)}{\partial T_s}, \tag{2.28}$$

$$[1 - D(.)]\left\{\frac{\partial G(T_g, n_i)}{\partial T_g}|_{n_i=n_i^*(T_g,T_s)} - \frac{\partial G(T_g, n_e)}{\partial T_g}|_{n_e=n_e^*(T_g)}\right\} = \frac{\partial C_g(T_g)}{\partial T_g}. \tag{2.29}$$

Where, by definition, $[1 - D(.)]$ takes a value from zero to one. Thus, at $T_s = 0$, the left hand side of (2.28) approaches infinity by Assumption 1. On the other hand,

[29]For details, see appendix.

at $T_s = 0$, the right hand side of (2.28) is zero from Assumption 2. But when T_s increases, the left hand side of (2.28) approaches zero from Assumption 1. This implies that there must be at least one $T_s > 0$ where the marginal value curve cuts marginal cost curve from above. This implies that the condition for Proposition 2, i.e. $T_s > 0$, holds so that $n_i^*(T_g, T_s) > n_e^*(T_g)$ is ensured. This in turn implies that the term in curly brackets in (2.29) is not equal to zero[30]. Given $n_i^*(T_g, T_s) > n_e^*(T_g)$, the left hand side of (2.29) approaches infinity at $T_g = 0$. But when T_g increases, the left hand side of (2.29) approaches zero from Assumption 1, once again. Thus, there must be at least one combination of $T_g > 0$ and $T_s > 0$ where the equilibrium condition for T_g and T_s is satisfied. From this, the following results emerge

Proposition 3. *Given that assumptions of the model hold;*

(i) With no/exogenous labour supply in the production function the firm optimal general training level T_g^* is zero (Becker's Results),

(ii) With labour supply in production but no specific training, i.e. $T_s = 0$, the firm optimal general training level T_g^* is zero (Becker's Results),

(iii) With labour supply in production and $T_s > 0$, the firm optimal general training level is positive, i.e. $T_g^* > 0$ (Violates Becker's Results),

(iv) Quit probability $D(W_E - C_e(n_e^*) + C_i(n_i^*) - W_2^*)$ is falling in general training (Violates Becker's Results),

(v) The firm optimal general training level $T_g^*(T_s)$ is increasing in the level of specific training T_s, and vice versa.

Proof. See the appendix. □

The intuition behind these results is clear. With no or exogenous labour supply and additive separability[31] of the production function in T_g and T_s, a given increase in general training increases the worker external value by the same amount as her value at the training firm $\frac{\partial G(T_g, n_i=1)}{\partial T_g} = \frac{\partial G(T_g, n_e=1)}{\partial T_g}$; the added effects of labour supply

[30]As will get clear in Proposition 3 below, this provides the bases for investment in general training.

[31]Note that assuming complementarity between T_g and T_s in the production function ensures all the above results without any labour supply term in the production function.

on general training's marginal productivity are missing now. In such a situation, the firm is not willing to invest because it cannot recover the training costs in period 2. The only way for the firm to recover training costs is not to pay the full marginal contribution of training to the worker. In other words, $\frac{\partial f(.)}{\partial T_g} > \frac{\partial W_2(.)}{\partial T_g}$. However, with no or exogenous labour supply and additive separability, the above conditions imply $\frac{\partial f(.)}{\partial T_g} = \frac{\partial W_E(.)}{\partial T_g} > \frac{\partial W_2(.)}{\partial T_g}$. But this will force the worker to leave and to get marginal product in the external market. The probability of stay only scales up or down the net marginal benefits of training once it is positive. In such a situation, the firm will invest in specific training only. Similarly, with no specific training, the marginal productivity of work is the same in training firm and the external market. This again leads to equal marginal productivity of general training in the training firm and the external market, so the firm has no incentives to invest in T_g. However, when $T_s > 0$ and labour supply is endogenous, it creates a wedge between the marginal productivities of general training in the training firm versus the external market even though T_g is fully transferable. The reason is that marginal productivity of work is high in the training firm due to the positive specific training. This forces the worker to work relatively more if she stays as in Proposition 2. But this in turn means that the marginal productivity of general training at the training firm is more than the marginal productivity of general training in the external market due to the complementary nature of work and skills in the production function. This wedge between marginal productivities of general training is the necessary condition for firm-financed general training as is highlighted in Section 2.2. Once this wedge exists, the fact that quit probability is falling in T_g further induces the firm to invest in general training.

Finally, as is clear from the proof of (v), the incentive complementarity between T_g and T_s arises from two channels. First, for a given gap between the marginal productivities of T_g at the training firm and the external market, more T_s decreases the probability of quit. This scales up the net marginal benefits from general training, the term $d(.)\{\frac{\partial G(T_g,n_i)}{\partial T_g}|_{n_i=n_i^*} - \frac{\partial G(T_g,n_e)}{\partial T_g}|_{n_e=n_e^*}\}\frac{\partial W_2^*}{\partial T_s}|_{n_i=n_i^*}$, and thus increases firm's incentives to invest in general training. Second, specific training increases work hours at the training firm as compare to the external market, which in turn increases the marginal productivity of general training because of complementarity between work and general training; the term $[1 - D(.)]\frac{\partial\left[\frac{\partial G(T_g,n_i)}{\partial T_g}|_{n_i=n_i^*}\right]}{\partial T_s}$. Because of these forces, general and specific trainings are incentive complements for the firm despite the fact that they are additively separable in the production function in (2.19).

2.4.4 Equilibrium Wage Effects of Trainings and Labour Supply

Now the study turns to the basic question of firm's rent charging behavior in equilibrium. For this purpose, it analyses the wage effects of trainings and labour supply. To see this, differentiate the wage function (2.22) at equilibrium values to get

$$\frac{\partial W_2^*}{\partial T_g} = \frac{\frac{\partial f(T_g,T_s,n_i^*)}{\partial T_g} + [1 + \frac{(1-D(.))\acute{d(.)}}{[d(.)]^2}][\frac{\partial W_E}{\partial T_g} - \frac{\partial C_e(n_e^*)}{\partial n_e}\frac{\partial n_e^*}{\partial T_g} + \frac{\partial C_i(n_i^*)}{\partial n_i}\frac{\partial n_i^*}{\partial T_g}]}{2 + \frac{(1-D(.))\acute{d(.)}}{[d(.)]^2}}, \tag{2.30}$$

$$\frac{\partial W_2^*}{\partial T_s} = \frac{\frac{\partial f(T_g,T_s,n_i^*)}{\partial T_s} + [1 + \frac{(1-D(.))\acute{d(.)}}{[d(.)]^2}][\frac{\partial C_i(n_i^*)}{\partial n_i}\frac{\partial n_i^*}{\partial T_s}]}{2 + \frac{(1-D(.))\acute{d(.)}}{[d(.)]^2}}, \tag{2.31}$$

$$\frac{\partial W_2^*}{\partial n_i} = \frac{\partial f(T_g,T_s,n_i^*)}{\partial n_i} + [1 + \frac{(1-D(.))\acute{d(.)}}{[d(.)]^2}][\frac{\partial C_i(n_i^*)}{\partial n_i} - \frac{\partial W_2^*}{\partial n_i}]. \tag{2.32}$$

Proposition 4. *Given Assumption 5 and optimal labour supplies;*

(i) Equilibrium wage is increasing in T_g, T_s and n_i,

(ii) It increases relatively more in T_g as compared to T_s,

(iii) The firm is charging rent on trainings, but it does not charge any rent on labour supply.

Proof. See the appendix. □

Point (iii) is discussed first as it is the main focus of this paper. As shown above, the firm is not charging any rent on working hours. The reason is that worker faces no uncertainty in period 2 and chooses labour supply where marginal cost of work is equal to its marginal wage benefits. This condition is the reason for why firm does not charge rent on labour supply. According to this, the firm optimal strategy would be to charge no rent on any factor that is decided by the worker in the model in period 2. On the other hand, any factor which is in the control of the firm, the firm charges rent on it. To see this clearly, it is better here to show that the firm will charge rent on capital as well besides rent on trainings when one introduces capital into the model. With capital k_i, the production function becomes $f(T_g, T_s, n_i, k_i)$. The firm second period wage setting strategy does not change with this. The only difference is that now the wage in (2.22) becomes a function of capital as well, as production and optimal labour supply become positive functions of capital. Then the derivative of wage function with capital becomes exactly as it is for the specific training in (2.31). This shows that the firm will pay back workers the value of capital that arises from complementarity between work and capital. But the firm charges rent on the direct contribution of capital into the production. Similar result will emerge if one introduces R&D into the model.

Regarding the wage changes with trainings, one can see from (2.30) that the equilibrium wage is increasing in general training marginal contribution to the training firm's production. Moreover, for $\frac{(1-D(.))d\acute{(.)}}{[d(.)]^2} > -1$, it also increases in general training effects on the worker external market value and dis-utility effects of labour when the worker works with the training firm; the last term in (2.30) shows this effect. On the other hand, wage decreases due to the dis-utility effects of training through work if she works in the external market. Regarding wage change due to specific training, wage is increasing in specific training contribution to the training firm marginal production and dis-utility effects of work. The external market factors do not enter the wage slope here because specific training is not transferable, and thus firm has no incentives to care about the external market when deciding wage package changes due to specific training. These results are quite plausible. In the case of general training, the firm needs to take into account external market forces when deciding on wage changes. It is needed to consider not only the change in marginal productivity of the worker in the internal versus the external market, but also the utility effects of labour in the two markets when deciding on wage changes. For example, if the labour cost functions are the same but very steep, then the extra work that the worker put when working with the training firm (remember $n_i^* > n_e^*$) will create greater dis-utility for the worker. Therefore, to make the worker put extra labour, the firm needs to increase the wage package accordingly. Moreover, for given productivities of trainings, the wage increase due to general training is more than the wage increase due to specific training. This is according to the theory of trainings and is plausible keeping the transferability of general training (Mincer, 1974b).

From the slope of wage function, one can see that the assumption of log concavity of $[1 - D(.)]$ is quite intuitive. If this assumption is violated, which would mean $\frac{(1-D(.))\acute{d(.)}}{[d(.)]^2} \leq -1$, the overall wage can still increase with training but the firm starts decreasing wage in the worker's outside option W_E. Nevertheless, this is counter intuitive and, therefore, I assume away such possibility in the model by imposing strict log concavity on $[1 - D(.)]$. For more details, without any assumption on $[1 - D(.)]$, first note that $\frac{(1-D(.))\acute{d(.)}}{[d(.)]^2}$ remains positive for critical values of quit below the mean, equal to zero at the mean and then negative for most of the distributions. This means that $\frac{(1-D(.))\acute{d(.)}}{[d(.)]^2}$ is decreasing in $D(.)$, becomes negative and then starts increasing until it reaches back to zero at $D(.) = 1$. In other words, the inverse hazard function $\frac{(1-D(.))}{[d(.)]}$ is a quadratic function in $D(.)$. Now, from (2.30) and (2.31) one can see that as $\frac{(1-D(.))\acute{d(.)}}{[d(.)]^2}$ tends toward zero, which means that $D(.)$ increases, the firm shares more and more of the marginal returns from trainings with the worker[32]. At $\frac{(1-D(.))\acute{d(.)}}{[d(.)]^2} = -1$, the wage change due to trainings is equal to the respective trainings marginal contribution to the worker value. What this mean

[32]To see this take values of one, zero and minus one, respectively, for $\frac{(1-D(.))\acute{d(.)}}{[d(.)]^2}$ and use $\frac{\partial W_2^*}{\partial n_i} = \frac{\partial C_i(n_i^*)}{\partial n_i}$ and $\frac{\partial W_E}{\partial n_e} = \frac{\partial C_e(n_e^*)}{\partial n_e}$.

is that when the probability of quit is increasing the firm increases the wage rate to counter the worker quit incentives. When $\frac{(1-D(.))\acute{d(.)}}{[d(.)]^2} = -1$ is reached, the firm end up offering all the direct value of training to the worker. If quit incentives still increase, i.e. $\frac{(1-D(.))\acute{d(.)}}{[d(.)]^2} < -1$, the firm starts decreasing wage in the worker outside option $(\frac{\partial W_E}{\partial T_g})$. Assumption 5 ensures that this scenario will not occur. Furthermore this assumption is widely used in the literature and is satisfied by many well known distributions[33].

One aspect of on-the-job training is the low training in the US as compared to countries like Germany and Japan. One reason, cited in the literature for this finding, is high turnover in the US economy. Acemoglu and Pischke (1998) provide specific examples within their setting to show that differences in quit rates across economies and multiple equilibria can explain such findings in non-competitive markets. In their model, high quit rate implies low monopsony power for the training firm and thus less incentive to invest in general training. Similar results can arise from this model as well. This result can be seen from the first order conditions in (2.28) and (2.29) and the findings that $\frac{\partial D(.)}{\partial T_g} < 0$ and $\frac{\partial D(.)}{\partial T_s} < 0$. Because of this, the left hand sides of (2.28) and (2.29) can increase in T_g and T_s in certain ranges and decrease at other values of T_g and T_s. This can lead to multiple equilibria. One can see this more clearly from the specific examples provided in the appendix.

The possibility of low training equilibrium is high for an economy for which the quit probability is high at zero level of training. Thus, economies where quit rate is very high may end up with low general and specific trainings, high quit rates and ultimately low labour supply. On the other hand, economies where quit rate is low initially, may never reach and trap in the low trainings equilibrium. In such economies, trainings and labour supply will be high and turnover will remain low. Which equilibrium will be efficient for the economy point of view? As is shown in the social planner problem below, the private trainings are below socially optimal

[33]Note from Equation (2.29) that Assumption 5 does not affect the criteria for investment in general training; it is only the positive gap in the curly brackets in (2.29) which matters for training investment. In the appendix, I provide two specific examples and its solution to develop understanding of the procedure used for getting the preceding results.

level of trainings. Second, the turnover in Equation (2.20) is inefficient as compared to the socially optimal turnover as is given in the social planner problem. Keeping this, the equilibrium with high trainings must be better for the society overall.

2.5 Extension and Discussion

This section addresses two questions. First, what will happen to firm's investment in general training if Assumption 6 is relaxed? Relaxing Assumption 6 means that the worker has to do two different tasks in the incumbent firm, with one needs only general skills and the other only specific skills. In such a case, the worker can decide how much time to allocate for working with the general skills task and how many hours to work with the specific skills task. It is shown below that the firm can still invest in general training if one also adds efforts into the model. But note once again that relaxing Assumption 6 has no effects on the firm rent charging behavior. The second question is about the comparison of the training results with the first best or socially optimal level of trainings. The section is concluded with a brief discussion of evidence on efforts and job-specific skills.

2.5.1 Endogenous Efforts along with Labour Supply

If a worker is allowed to choose time allocation between working with general and specific skills, then the results of Proposition 3 can change. This question is important when both specific and general skills are assumed to exist and additively separable. For example, if the worker spends less time on general tasks in the training firm as compared to the external market, the results of positive investment in general training are no longer guaranteed. Note that in practice, such a distinction may not be possible mostly but this is an important question in theory at least. It is shown below that adding efforts, in addition of working hours, still guarantee Proposition 3. Efforts are those features of the labour contract that enhance productivity but workers dislike it as in Leamer (1999). Unlike the trade-off in working hours, putting more effort with $G(.)$ does not stop the worker from working sincerely with $S(.)$. Given that the worker has no preferences over $G(.)$ and $S(.)$, one cannot expect the worker to put different efforts while working in two different tasks within the same firm.

Suppose that the production function in (2.19) is now $f(T_g, T_s, n_g, n_s, e_i) = G(T_g, n_g, e_i) + S(T_s, n_s, e_i)$, where n_g denotes hours spent working in tasks that need general skills, n_s denotes hours spent working in tasks that need specific skills and e_i denotes efforts in the incumbent firm. If the worker leaves the training firm, her value will be $f(T_g, 0, n_e, 0, e_e) = G(T_g, n_e, e_e)$. Efforts costs are denoted by $C(e_i)$ and $C(e_e)$. Suppose the production and efforts cost functions satisfy Assumptions 1-3. Additionally, now utility is assumed to be linear in leisure l_i or l_e and time allocation is as $M = l_i + n_g + n_s$ or $M = l_e + n_e$, depending on where the worker works. Equation (2.20) now becomes

$$W_2 + \theta - C(e_i) + l_i < W_E - C(e_e) + l_e. \tag{2.33}$$

Now, if the worker stays, she has to decide the allocation of time between specific and general tasks in addition of efforts choice. The firm's problem is not changing at all; it has to offer period 2 wage and decides on trainings in period 1. Once W_2^* and W_E offers are there, the worker is in a position to decide about quit, allocation of time and efforts. The first-order conditions of the worker decision are

$$\begin{gathered}
\frac{\partial G(T_g, n_g, e_i)}{\partial e_i} + \frac{\partial S(T_s, n_s, e_i)}{\partial e_i} = \frac{\partial C(e_i)}{\partial e_i}, \\
\frac{\partial G(T_g, n_e, e_e)}{\partial e_e} = \frac{\partial C(e_e)}{\partial e_e}, \\
\frac{\partial S(T_s, n_s, e_i)}{\partial n_s} = 1, \\
\frac{\partial G(T_g, n_g, e_i)}{\partial n_g} = 1, \\
\frac{\partial G(T_g, n_e, e_e)}{\partial n_e} = 1.
\end{gathered} \tag{2.34}$$

The first two equations help decide efforts in the two markets whereas the last three equations determine time allocation. The third and fourth equations in (2.34) can be used to determine time allocation between specific and general skills tasks if the worker decides to stay. From this two equations one can see that the worker would like to allocate more time to the tasks where she has more skills. Furthermore, one can get the following result

Proposition 5. *Given the production and cost functions above, $n_g^* > n_e^*$, $n_s^* + n_g^* > n_e^*$, and $e_i^* > e_e^*$ must hold.*

Proof. The proof follows from the above equations. The equality of marginal productions of n_g and n_e in the last two equations in (2.34) implies that $\frac{\partial G(T_g,n_g,e_i)}{\partial e_i} = \frac{\partial G(T_g,n_e,e_e)}{\partial e_e}$ must hold. Given the assumptions in the production function, this then implies that $\frac{\partial C(e_i)}{\partial e_i} > \frac{\partial C(e_e)}{\partial e_e}$ from the first two equations of (2.34). But from Assumption 2, this gives $e_i^* > e_e^*$. However, for $e_i^* > e_e^*$, the equality of marginal productions of n_g and n_e in the last two equations in (2.34) can hold only if $n_g^* > n_e^*$. Then, $n_s^* + n_g^* > n_e^*$ must hold as n_s^* cannot be negative under the assumptions on the production and cost functions. □

This result implies that endogenous efforts can further the firm incentives to invest in general training. The presence of $n_g^* > n_e^*$ and $e_i^* > e_e^*$ mean that the wedge in the productivities of general training in the training firm versus the external market in Equation (2.29) should be even bigger. The third and fourth equations in (2.34) imply that, given the production functions for general and specific skills, the worker will allocate more time to the task where she has more skills. Note that in getting this result, the convexity of effort cost function and linearity of utility in leisure play an important role. The convexity of effort cost function is most plausible and widely used specification in the efficiency wage literature. Risk neutral utility is widely used at least in theory. If one relaxes this assumption, it does not mean that the above result will not hold. The only thing is that under convex utility cost function in labour, many possibilities may arise. In such a case, one is required to rely on specific functional specifications. Importantly, the result about the effects of working hours on wage is the same as earlier. To see this, the wage slope will now look as

$$\frac{\partial W_2^*}{\partial n_g} = \frac{\partial f(T_g, T_s, n_g^*, n_s^*, e_i^*)}{\partial n_g} + [1 + \frac{(1 - D(.))\acute{d(.)}}{[d(.)]^2}][1 - \frac{\partial W_2^*}{\partial n_g}].$$

But the second term cancels out due to the first order-condition for optimum n_g. Thus, the worker gets full marginal contribution of an work hour, given that other things remain the same.

2.5.2 Comparison with Social Planner Choices

This section highlights how the individual firm choices can be different from social planner's solution. From the above analysis, one can see that the efforts and labour choices involve no inefficiencies. The reason is that efforts and work hours choices are made in period 2, where the worker has already taken the quit decision. Thus efforts and work hours choices involve no uncertainty, and the worker decides its levels by equating marginal cost to marginal return in each market. Thus, this section ignores efforts for simplicity for this analysis.

The problem before the social planner is training decision. The social planner does not know the utility shock at the beginning of period 1, where it has to decide on trainings. Thus, the planner does not know whether a typical worker will stay in the training firm or will leave in period 2. Moreover, the social planner is not supposed to affect the quit decision by offering incentives. However, the training decision can have effects on quit probability. For social planner, the efficient quit would be $f(T_g, T_s, n_i) + \theta - C_i(n_i) < G(T_g, n_e) - C_e(n_e)$. Thus, the quit probability becomes $D(G(T_g, n_e) - C_e(n_e) - f(T_g, T_s, n_i) + C_i(n_i))$. Now, here arises the important difference. In a model with additively separable trainings but no labour, the planner is not supposed to worry about the effects of quit on general training as it is equally productive in every firm. Thus, the planner has to consider the possibility of quit while deciding on specific training only. But with labour supply in production, the planner needs to worry about the effects of quit on general training as well. The reason is that different levels of labour supplies create differences in the productivity of general training in the training firm versus the external market. So, the social planner decisions of both general and specific trainings involve uncertainty. The social planner's problem thus can be formulated as;

$$(T_g^s, T_s^s) = \underset{\{T_g, T_s\}}{arg.max} \left[\{1 - D(..)\} \left[G(T_g, n_i) + S(T_s, n_i) \right] + D(..) G(T_g, n_e) - C_s(T_s) - C_g(T_g) \right], \tag{2.35}$$

where $D(..)$ is short hand for $D(G(T_g, n_e) - C_e(n_e) - f(T_g, T_s, n_i) + C_i(n_i))$. One can see from above that the worth of general training is $[1 - D(..)]G(T_g, n_i) + D(..)G(T_g, n_e)$ for the social planner versus its worth of $[1 - D(.)]G(T_g, n_i)$ for the private firm. From this, the following first order conditions emerge

$$\begin{gathered} \frac{\partial G(T_g, n_i)}{\partial T_g}[1 - D(..)] + D(..)\frac{\partial G(T_g, n_e)}{\partial T_g} = \\ d(..) \left[G(T_g, n_i) + S(T_s, n_i) - G(T_g, n_e) \right] \left[\frac{\partial G(T_g, n_e)}{\partial T_g} - \frac{\partial G(T_g, n_i)}{\partial T_g} \right] + \frac{\partial C_g(T_g)}{\partial T_g}, \end{gathered} \tag{2.36}$$

$$\frac{\partial S(T_s, n_i)}{\partial T_s}[1 - D(..)] + d(..)[\frac{\partial S(T_s, n_i)}{\partial T_s}]S(T_s, n_i) = \frac{\partial C_s(T_s)}{\partial T_s}. \tag{2.37}$$

The first term on the left hand side in (2.36) shows the marginal contribution of general training to the output if the worker stays, whereas the second term is the marginal contribution of general training if the worker leaves. The first two terms in the bracket on the right hand side show the benefits of training in terms of increasing probability of stay due to general training. Both this terms increase the social planner incentives to invest in general training as general training is more productive at the training firm. The last bracketed term shows the decrease in stay probability due to general training effects on the worker outside value. In (2.37), the first term shows the marginal contribution of specific training to the training firm output whereas the second term shows the effects through increase in probability of stay due to specific training.

Proposition 6. *(i) The socially optimal trainings must be greater than privately decided trainings, i.e.* $T_g^* < T_g^s$ *and* $T_s^* < T_s^s$*,*

(ii) The private general training choice is more distorted than specific training decision.

This can be seen by comparing (2.36) and (2.37) with (2.28) and (2.29). In the case of general training, the social planner keep in views the benefits to both the training firm and the external market, whereas the private training firm only exploits the difference between the values of general training in the two markets; the curly bracket term in (2.29), for a given quit probability. Not only this, the social planner's training decision increases the probability of stay as shown in the first term on the right hand side of (2.37). For the private firm, training increases the probability of stay, but it is offset by the fact that the training firm takes into account the effects of increasing training on wage; see Equation (2.27). For the specific training, the only inefficiency is one that arises due to this last factor. That is, the increase in probability of stay due to specific training is offset by the increase in wage due to specific training in case of private training decision. Taking this into account, the training firm invests relatively less in specific training as compared to the social planner; compare Equation (2.28) to (2.37).

In the case of general training, three types of distortions exist. One is that the private firm ignores the benefits of training to the external market, $D(..)G(T_g, n_e)$. Second, the increase in probability of stay is offset by the training firm requirement of increasing wage. Third, the training firm invests upto the point where the marginal productivity of general training is above its marginal productivity in the external market; the term in the curly brackets in (2.29). Thus, although general training in the market economy is not zero as in Becker's results, but there is a substantial under investment in general training by the market economy.

Before conclusion, evidence on efforts and specific skills are briefly discussed here. Research in human resource management shows that workers training leads to job satisfaction, commitment with the organization and high quality service delivery (Bartlett, 2001, Rowden and Conine Jr, 2005, Schmidt, 2007, Costen and Salazar, 2011, José Chambel and Sobral, 2011). In labour research also, evidence shows that efficiency wage, gift exchange and reciprocity considerations are important in the real world and that efforts matter and are increasing in monetary and non-monetary benefits (Akerlof and Yellen, 1990, Peach and Stanley, 2009, Kube et al., 2012, Cohn et al., 2014). Moreover, studies show that efforts can increase also for a given wage because of reciprocity or gift exchange consideration (Rotemberg, 2006, Dohmen et al., 2009). Field experiments have clearly revealed that reciprocal workers get more training and that workers reciprocate firm investment in training by exerting more efforts (Leuven et al., 2005, Englmaier et al., 2015, Sauermann, 2015). Additionally, firm-specific skills existence and importance is always a point of interest for researchers. Its estimation is a complicated problem due to many conceptual and estimation problems. Overall, empirical estimates on the tenure-wage relation are consistent with the existence of firm-specific skills[34]. Similarly, Kessler and Lülfesmann (2006) provide examples of compelling evidence on the existence of firm-specific skills.

2.6 Conclusion

This study reports evidence from existing studies which shows that firms are charging rent on capital, training and R&D investment. However, the result about whether firms charge rent on working hours is ambiguous in the existing empirical literature. The study further investigates this by using firm-level panel data of Belgium for the period 1997-2006 from Konings and Vanormelingen (2015). The results from these estimations consistently show that firms are charging rent on

[34]For detail discussion, see the notes of Acemoglu and Autor on: https://economics.mit.edu/files/4689.

capital and training but they do not charge rent on working hours. The increase in wage from an additional work hour is at least as high as the corresponding increase in firm's value added from that additional work hour, after controlling for capital, training, contract type, schooling and fixed firm effects. This result holds for different specifications, and after controlling for possible endogeneity of training, work and capital.

The study shows that if wage is set through bargaining process and/or the worker's quit decision is exogenous then the existing non-competitive models of human capital accumulation through training cannot explain this empirical finding under standard production functions; where inputs like capital, working hours, human capital etc. are assumed to be technological complements. To explain the above evidence better, this study develops a model where the firm invests in both firm-specific and general training and the worker takes optimal quit decision. It shows that in such a setting the firm's optimal wage strategy is to charge rent on any factor which is under the firm's control; like training and capital. On the other hand, the firm is charging no rent on factors where the worker decides its supply like work hours, job efforts etc. The study also presents evidence which is indicative of the possibility that the joint existence of firm-specific and general transferable skills can better explain firm's investment in workers general skills as compared to explanations that are based on asymmetric information. The study also contributes to the theory of training by showing one novel way of why firm can finance general training.

These findings are important in many aspects. Wage bargaining process such as Nash bargaining is widely used in labour economics as a tool of wage determination in an imperfectly competitive environment. Similarly, the assumption that worker's time and her skills are complementary with each other and with other inputs is widely used, accepted and is a plausible assumption. But this together cannot withstand the empirical finding that firm does not charge rent on work hours. Additionally, many theoretical explanations, particularly asymmetric information and job-specific skills, are put forward to support the evidence of firm-financed general training. But there are no formal evidence of which explanation is a likely reason of such investment. Acemoglu and Pischke (1998) emphasize asymmetric information about workers skills whereas Kessler and Lülfesmann (2006) put some evidence in favor of firm-specific skills. This study presents more evidence using Belgium firms which supports the existence of job-specific skills.

Chapter 3

Schooling, Job Training and Wage Inequality Amplification

3.1 Introduction

During the 1990s and the 2000s, many empirical studies of job related training consistently confirm that workers who enter jobs with a relatively high pre-job schooling also get more on and off-the-job training in the US, UK and Europe (Altonji and Spletzer, 1991, Booth, 1991, OECD, 1991, 1999, 2003, Green, 1993, Veum, 1997, Bartel and Sicherman, 1998, Arulampalam and Booth, 2001, Brunello, 2001, Arulampalam et al., 2004, Verhaest and Omey, 2013). Interestingly, the evidence suggests that the different training rates among workers with different schooling levels is mainly due to differences in workers' willingness to train rather than firm's decisions in training provision (OECD, 1999, Leuven and Oosterbeek, 2000, Bassanini and Ok, 2004, Maximiano and Oosterbeek, 2007, Pew Research Center, 2016). One of the questions it raises is why and under what conditions the highly educated workers would like to take relatively more job training. The second interest it creates is its wage implications, as the positive association between schooling[1] and job training implies that pre-job skills heterogeneity is amplified during job on the one hand and

[1]Hereafter, this study uses the term schooling for pre-job formal education.

the worker may be needed to pay all or part of the training costs on the other hand. Additionally, the question of whether the worker spares time for training by cutting down working hours or her leisure time is important for wage inequality. This ultimately implies that, for a given aggregate market demand of skills, wage inequality can change significantly in the post-entry period of job depending on the quantity of training that an individual obtains and the time allocation between training, work and leisure.

Considering these questions, the current study develops a model where workers invest both goods and time in skills accumulation[2]. It uses a representative firm for the analysis and assumes that the time and goods investment in human capital leads to completely general skills. The assumption of general skills from investment in training is consistent with the empirical evidence that most of the trainings in the real world are general in nature (Loewenstein and Spletzer, 1999). In the model economy, there are no frictions, uncertainty or asymmetric information. So, the worker is assumed to pay all the costs for such training. The notion that workers should pay for general training is the building block of human capital theory of training in competitive markets (Becker, 1962, Mincer, 1974a, Waldman, 1984, Acemoglu and Pischke, 1998, 1999, Booth and Zoega, 2004, Friedrich, 2016). In the model, schooling can make worker's training more productive in human capital accumulation[3], can affect the worker's patience level, change their preferences for leisure, or can affect their inter-temporal elasticity of substitution[4]. Moreover, labour supply is endogenous in the model and the worker allocates time between work, time in training and leisure.

Given this setup, the study shows that individuals' time investment in training can increase in their pre-job schooling even if utility is decreasing in training time and high-skilled individuals pay more than low-skilled individuals for a given training

[2]Individuals can accumulate human capital by investing time in learning such as getting on-the-job and off-the-job training. They can accumulate human capital by investing in goods such as laptop and in health even. See Section 3.2 for more details.

[3]This is called direct effects or human capital efficiency effects of schooling.

[4]See for example, Heckman et al. (1998), Fouarge et al. (2013), Havranek et al. (2015) etc.

time, given that schooling either improve the efficiency of training time in human capital accumulation, or affect worker's preferences. But all the possible preference effects of schooling are on their own not enough to generate increasing human capital and per hour gross wage functions in schooling. On the other hand, in the case of direct human capital effects of schooling, net of investment costs wage function, consumption and human capital increase in pre-job schooling. The generated wage function can at least partly match US median earning distribution by schooling categories. Interestingly, in such a specification, working time decreases but leisure time is independent of schooling level of individuals. This means that the workers are utilizing the time they get from working less in learning rather than in leisure. Thus, the specification which can generate realistic wage, human capital and consumption functions in pre-job schooling is one where pre-job schooling improves individual's efficiency of human capital accumulation.

This study contributes in many aspects to the existing literature on human capital accumulation and inequality. In typical human capital accumulation models, the only cost of skills accumulation is the opportunity cost that arises since the available time can either be used for production work or for human capital accumulation (Lucas, 1988, Rebelo, 1991, Heckman et al., 1998). In this study, in addition to costs due to such a trade-off, the worker is needed to bear additional direct cost of training[5]. This study shows that the assumption that pre-job schooling has efficiency effects in human capital production is enough to not only create positive schooling-training association under many different formulation but wage distribution that seems more realistic. Moreover, it can generate the empirical findings that preferences for leisure, future orientation etc. are the factors behind positive association between schooling and training (Fouarge et al., 2013). But the model predicts that these factors might be of secondary importance.

[5]This cost can arise as tuition fee of off-the-job training, salary of trainer, rent of training centre, cost of training materials etc. This is what makes training different from a typical human capital accumulation story.

Most importantly, in typical macroeconomics models time is split into labour-leisure or production-human capital accumulation only, besides the assumption that skills are firm-specific in many cases (Lucas, 1988, Rebelo, 1991, Heckman et al., 1998). This study allows time allocation into work, training and leisure. The prediction that highly qualified individuals optimally work for less time, but they do not spend more time in leisure is new and interesting. On one hand, it implies that studies in this literature that assume exogenous labour supply may over estimate the total wage effects of differences in schooling, abilities etc., as low schooling/abilities workers work for more time under endogenous labour supply (Heckman et al., 1998). On the other hand, specifications with only labour-leisure choices may not be rich enough, as in such studies low work time by definition implies more leisure time. But before studying the implications of this prediction, it would be ideal to empirically investigate the association between schooling and leisure. Many household surveys ask questions about leisure activities, so one can use it to precisely define leisure time and study how it changes with qualifications.

After the introduction, the theory of job training and school-training association are briefly discussed in Section 3.2. Section 3.3 contains the model and its results. Section 3.4 concludes the study.

3.2 Theory of Training and the Empirics on School-Training Association

Before delving into discussing school-training association, it is worthy to discuss the existing theories of training in order to develop an understanding of training literature. The firm or worker can affect working skills in a number of ways: through formal on-the-job training, off-the-job instruction seminars, vocational and technical training, business school, apprenticeship, correspondence courses, overall healthy working environment or going for further formal studies. The famous historical apprenticeship program is a classic example of job related training. The important thing about training which makes it different from schooling is that the worker is getting it while being in an employment relationship and is likely to receive wage even if on training rather than routine work. Moreover, training is called general if it leads to skills which are effective in number of firms, and firm-specific if it has led to skills that are useful only in the incumbent firm.

The job training literature begins with Becker's (1962) study which shows that a competitive firm will not pay for general training of workers because general training has the same characteristics as technology. If the firm tries to recover the costs of training by paying lower wages in post-training period, the worker will quit to get according to her marginal product in a frictionless market. The conclusion of Becker's (1962), also see Mincer (1974a) and Becker (1975), study is that general training will improve productivity and will lead to an equivalent wage increase for the workers in competitive markets, so the worker should pay the cost of general training. On the other hand, the firm can pay for a firm-specific training, where it can recover the cost of training by paying wages lower than the marginal product to trained workers in post-training period. The reason for the firm monopsony power in case of firm-specific training is that workers lose the skills if they leave the parent firm (Becker, 1962, 1975, Mincer, 1974a, Waldman, 1984, 1990, Acemoglu and Pischke, 1998, Friedrich, 2016). The first point of this theory is that training should lower firm's output during current period when worker is getting training instead of doing the routine production work. Secondly, it predicts that workers should finance general training in perfectly competitive markets with no mobility barriers. Thirdly, the wage increase due to general training should be higher than the wage increase from specific training (Acemoglu and Pischke, 1998, Frazis and Spletzer, 2005).

Practically, every training program run or suggested by a firm for its workers is likely to have both general and specific aspects to a certain extent. But this distinction is important for research because only then one can investigate the prediction that workers should pay for general training. To this end, studies have used different indirect approaches to determine about the nature of training. For example, training is defined as general when it increases wage at the current as well as future employers whereas the one as specific which increases wage only at the current employer (Frazis and Spletzer, 2005). Similarly, training which reduces mobility of workers is also defined as specific training as predicted by human capital theory. The empirical studies that check these predictions gave mixed results, particularly about the prediction that firms cannot finance general training. Many studies are suggestive of the fact that firms are at least partially financing seemingly general training (Acemoglu and Pischke, 1998, Barron et al., 1999, Loewenstein and Spletzer, 1999). On the other hand, some studies support the standard competitive theory of training (Veum, 1995b, Sousounis, 2009, Jones et al., 2011).

Keeping the lack of support on wage compression in the case of general training, particularly in the studies that were based on the National Longitudinal Survey of Youth (NLSY) during the 1990s, different studies introduced market imperfections into the Becker's (1962) model of training investment. Particularly, assumptions such as asymmetric information and mobility barriers were introduced to answer the question of why should firm pay for general training of workers (Katz and Ziderman, 1990, Acemoglu and Pischke, 1998, 1999, Loewenstein and Spletzer, 1999, Booth and Zoega, 2004). Thus, the focus of the theoretical literature of training after Becker (1962) is to determine that whether positive level of general training will take place in equilibrium, and if yes, the firm or the worker will pay for it. The question of what type of workers would like to get training and its wage implication is relatively under investigated theoretically.

Regarding implications of training, particularly for inequality, one important question is whether the more educated or less educated workers get on and off-the-job training[6]. Though the theoretical literature is silent on this question, the empirical research has considered questions such as; whether male workers get more training than females, workers in which professions or positions usually get more training

[6]This question is of critical importance for policy purposes because if more educated people get more training, although it can enhance growth, it will probably lead to more wage inequality.

and whether highly educated or low educated people get more training once they join jobs. The first study related to this literature is Lillard and Tan (1986). It uses Current Population Survey (CPS), three cohorts of National Longitudinal Survey (NLS) and Employment Opportunities Pilot Survey (EOPP) to find that formal schooling and on-the-job training are strongly complementary in the US. Barron et al. (1989) use a unique data sponsored by the National Institute of Education and the National Center for Research in Vocational Education (1982) and find that those firms select high ability workers in jobs which give more on-the-job training. But this paper is silent on whether high ability workers want to get more training or firms train high ability workers. Altonji and Spletzer (1991) use NLS of the high school class of 1972 and the Dictionary of Occupational Titles and find that post-secondary education and aptitude has strong positive effects on workers training. They find that major part of the positive effects of post-secondary education on training is explained by high achievements at the end of high school.

Using British Social Attitude Survey (BSAS) of 1987, Booth (1991) establishes a strong complementarity between schooling and formal vocational training. Lynch (1992) study shows that schooling increases the likelihood of off-the-job training and apprenticeship but its impact on on-the-job training is smaller. Green (1993) uses data on 7,969 employees from General Household Survey (GHS) of 1987 and confirms the earlier result that higher qualifications: A level, O level, higher degree, vocational and other qualifications have a strong positive relation with job related training participation. He also finds that recent entrants to jobs and high status jobs need more training participation. The estimates on the relation of different qualifications with length of trainings by male and females give mixed results in this study. Veum (1997) uses NLSY data between 1987-92 and finds that on and off-the-job company financed training is positively correlated with education, ability and prior tenure on the job. Employee financed training has positive but insignificant relation with education. Bartel and Sicherman (1998) draw almost similar results on the company training as that of Veum (1997) from NLSY between 1987-92.

Regarding studies in the 2000s, Arulampalam and Booth (2001) study, based on National Child Development Survey (NCDS) in Britain, also finds positive and significant relation between training and degree, A level and O level qualifications. The relation between training and vocational qualification, and training and apprenticeship is positive but insignificant. In their study, the dependent variable training is measured, alternatively, as training incidence and the number of training courses taken between 1981-91 period. Degree and A level qualifications has the highest coefficient in education-training relation. Brunello (2001) uses the European Community Household Panel (ECHP) data for 1994-96 period to study education-training complementarity in European countries. According to his results, individuals with college degree have highest probability of investing in further skills in Europe, followed by individuals with upper secondary education. Individuals with lower secondary education have the lowest probability to invest in training. The only country in which higher education reduces the incidence of training is the Netherlands. In this study, educational qualification has no significant effects on the duration of training. Arulampalam et al. (2004) use the ECHP panel between 1994-99 to study the factors behind training investment in ten European countries. They show that in nine countries, workers with tertiary education are more likely to get training as compared to workers with less than upper secondary education. Finally, the positive education-training association is also reported in various reports by Organization of Economic Co-operation and Development (OECD, 1991, 1999, 2003).

As mentioned earlier, this positive association between education and job training can have important policy implications regarding wage inequality. If this relation is because of firms' characteristics, as many evidence suggest that high technology and greater size firms hire better workers and supply more training, then it may call for different policy options than if it is due to differences in workers' choices. Particularly, the OECD reports and many other studies suggest that this relation is mostly because of differences in the workers' demand for training[7]. In any case, given the

[7]See for example, Leuven and Oosterbeek (2000), Bassanini and Ok (2004) and Maximiano and Oosterbeek (2007).

increasing importance of skills caused by technological innovations, firms and workers are likely to rely on more and more job training and this will further increase the pre-job skills heterogeneity when more educated workers get more training. For example, Arulampalam et al. (2004) show that each year more than one-third individuals are likely to start training in Britain, Denmark and Finland. On the other hand, in Ireland, Italy, and the Netherlands, less than 10% start training each year. In Austria, Belgium, France, and Spain the proportion of workers getting training ranges from 10% to 16%. This can lead to ever increasing wage inequalities among the different schooling groups. Given this discussion, in the next section the study tries to find the conditions under which highly qualified workers would like to get more training during jobs and its wage effects.

3.3 Investment in Training and its Wage Effects

This section pursues the question of why highly qualified individuals would like to invest more time and resources in human capital accumulation and its wage implication. This question is pursued in a standard neoclassical infinitely lived growth model, as most of the human capital accumulation models follow this path. This will help in studying the effects of differences in pre-job schooling on the balance growth path. The possible effects of pre-job schooling during job can be its effects on entry level human capital, it can enter the human capital production function, can affect patience level of workers, can influence inter-temporal elasticity of substitution or can have effects on worker's preferences about labour-leisure choices. These possible effects would be considered one after the other to see its implications for school-training association and wage inequality. This model is more general and the individual in the model can also be considered as household if it is desired to study the effects of differences in family income or education on future generations skills and wages or mobility along the income ladder.

The study starts with a closed economy populated by N individuals. Differences in schooling levels is the source of making workers ex ante heterogeneous, and this will be captured by schooling effects on human capital accumulation efficiency, patience level differences etc. Furthermore, future human capital production is assumed to depend on existing stock of human capital, spill-over effects from the firm and time and goods investment in human capital. Firms are assumed equal in every aspect. Therefore, a representative firm is used for the analysis. There is a single good in the economy and its price is normalized to one. Note that the notion that different firms produce homogeneous good with price normalized to one is standard in growth literature (Aghion et al., 2002). This assumption is made to abstract from the demand side effects of relative prices. Thus, there is only consumer heterogeneity in this model. Finally, in the literature of on-the-job training, credit constraints play an important role because in credit constrained economy training is not possible if it makes the during training wage negative (Acemoglu and Pischke, 1998, 1999, Booth and Zoega, 2004). In this model no such a constraint is in place[8]. Given this, it is shown below that under certain conditions, workers who are highly educated strictly invest more in training and resultantly their level of human capital and wages are increasing in pre-job schooling in the long run.

3.3.1 Production

Worker's good investment in human capital does not concern the firm as its akin to investing in health or laptop etc., which is a worker's private decision. Similarly, when training time T_{it} is assumed to lead to general skills and worker is paying for it, then the firm role becomes passive in training as well. On the other hand, if training is firm-specific or a firm is sharing cost of general training with the worker, then the firm must play an active role to determine the level of training (Becker, 1981). Given the evidence that most of the skills gotten through training are general, it is

[8] In this model, workers can pay from their profit income when net of investment costs wage is negative.

assumed that time investment in training leads to general skills and workers pay all the costs of training as predicted in the Becker's competitive theory of training. This also helps prevent the firm's role in training, so that one can focus on differences in workers' choices. The assumption that the worker pays all training costs also helps one get conservative results about the wage effects of training. Thus, the firm's job is to produce the final good by hiring effective human capital services in the model. The study starts with the following production function

$$Y_t = A_t^{1-\beta} Q_t^{\beta}, \tag{3.1}$$

where $Q_t \equiv \sum_{i=1}^{N} h_{it} x_{it}$. Y_t is output of the firm at time t, A_t is the available technology in the economy at time t, h_{it} is individual's i human capital at time t, x_{it} is individual's time spent in work. The term Q_t denotes the effective labour hours that the firm chooses in each period. Given the assumption that workers will bear the full costs of training, the choice of T_{it} is irrelevant to the firm's problem. Note that the firm only cares about (i) how much time workers spent on work x_{it} and (ii) their human capital h_{it} which embodies all the knowledge generated by pre-job schooling and previous trainings. If workers choose to spend more time on training (i.e., T_{it} increases), this will be reflected in the labour market as a reduction in aggregate labour supply $\sum_{i=1}^{N} h_{it} x_{it}$. Formulated in this way, the production side problem is standard and really simple. The firm's maximisation problem is given by

$$\Pi_t = \underset{Q_t}{max} \left\{ A_t^{1-\beta} Q_t^{\beta} - w_t Q_t \right\}, \tag{3.2}$$

where w_t is the market wage rate per unit of effective human capital Q_t. The first-order condition of the firm's problem is

$$w_t = \beta A_t^{1-\beta} Q_t^{\beta-1}. \tag{3.3}$$

Using this we can get

$$Y_t = A_t \left(\frac{\beta}{w_t} \right)^{\frac{\beta}{1-\beta}} \quad \text{and} \quad \Pi_t = (1-\beta) Y_t. \tag{3.4}$$

3.3.2 Consumption and Human Capital Investment

Pre-job schooling differences can affect workers in many ways. It can result in human capital differences at the time of entry into the labour market[9], can affect worker's ability of human capital accumulation, behavior about future discounting, individuals' inter-temporal elasticity of substitution or their preferences for leisure (Heckman et al., 1998, Aloi and Tournemaine, 2013, Fouarge et al., 2013, Havranek et al., 2015). The impacts of all these possible schooling effects on training and wages are checked. With this, the infinitely lived consumer i maximises the following utility function in this model economy

$$maxU_i = \sum_{t=0}^{\infty} \phi_i^t \frac{[c_{it}^{\rho_i} l_{it}^{1-\rho_i}]^{1-\sigma_i}}{1-\sigma_i}, \tag{3.5}$$

subject to

$$c_{it} = w_t h_{it} x_{it} - h_{it} T_{it} - I_{it} + \frac{1}{N} \Pi_t, \tag{3.6}$$

$$h_{it+1} = S_i \left(T_{it} h_{it} \right)^{\upsilon} \left(\frac{Y_t}{N} \right)^{1-\upsilon} + \varphi I_{it} - \delta h_{it}, \tag{3.7}$$

$$T_{it} + x_{it} + l_{it} = 1, \tag{3.8}$$

$$T_{it}, c_{it}, x_{it}, I_{it} \geq 0, \tag{3.9}$$

and the initial conditions $h_{i0} > 0$,

where $\phi_i, \varphi, \upsilon, \rho_i, \delta \in (0,1)$ and $\sigma_i > 0$. Here ϕ_i is the individual i discount factor, ρ_i is the share of consumption in the utility of individual i and σ_i is the inverse

[9] This can be captured by initial human capital differences in this model.

of inter-temporal elasticity of substitution. Schooling S_i is assumed to affect discount factor $\phi(S_i)$, preference between consumption and leisure $\rho(S_i)$, inter-temporal elasticity of substitution $\sigma(S_i)$ and enters the human capital production function directly as in the above specification. But to save space, the study uses ρ_i instead of $\rho(S_i)$ and so on. Moreover, υ is the combine share of training time and human capital or effective training whereas $(1 - \upsilon)$ is the share of the spill-over effects in human capital production, δ is the common depreciation rate of human capital, c_{it} denotes consumption of individual i, l_{it} denotes individual's leisure time, T_{it} denotes the time that individual i spends in job training, x_{it} shows their working time and I_{it} denotes the investment of goods in human capital accumulation with its per unit productivity of φ. S_i is individual's exogenous level of schooling and h_{it} is their human capital at time t. The profit of the firm Π_t is equally distributed among the N individuals. As in Ben-Porath (1967) and Heckman (1976), human capital is embodied in the worker and is productive both in the final good production and its own production. Also note that individual's future human capital accumulation depends on their exogenous schooling level S_i, time investment in training T_{it}, her current human capital level h_{it}, good investment I_{it} and spill-over effects from the firm $\frac{Y_t}{N}$.

One can see that the $h_{it}T_{it}$ term is not appearing in the budget constraints of standard human capital models. Such models carry the opportunity cost of time allocation to human capital production as the only cost of skills accumulation. But in most instances, workers or firms are supposed to pay direct costs of training as well. For example, Employer Skills Survey 2017 shows that 66% of the establishments trained their staff in the UK over the last 12 months incurring training expenditure of £44.2 billion (Winterbotham et al., 2018). 62% of the staff were trained and the expenditure per trained employee is £2,470. Similarly, when workers are getting off-the-job training like vocational schools, technical institutes training or further studies they will have to pay tuition fee besides the forgone income from not working. Given the assumption that workers are paying all the costs of training, this direct cost will be captured by $h_{it}T_{it}$ in the net wage $w_t h_{it} x_{it} - h_{it}T_{it}$. Another important feature of (3.6) is that to get T_{it} training, the worker is paying $h_{it}T_{it}$. This means that highly skilled workers are paying relatively more for a given time in training. This is introduced to capture the fact that highly qualified are likely to spend T_{it} time in more advance courses that are likely to be more expensive as compared to courses for low skilled workers. Finally, note that spending of $h_{it}T_{it}$ enters into human capital production as T_{it}^{v} only. The term h_{it}^{v} in (3.7) represents contribution of existing human capital in new human capital production.

The human capital accumulation equation of individual i is standard except the assumption of additive separability between good investment in human capital and effective training, and a little different way of introducing spill-over effects $\frac{Y_t}{N}$ (Heckman, 1976, Rebelo, 1991, Heckman et al., 1998, Munandar, 2008, Aloi and Tournemaine, 2013). The exogenous schooling level S_i increases the efficiency of effective training in future capital generation directly besides its possible effects on ϕ, ρ and σ. The assumption that schooling increases the efficiency of training in future capital generation is more plausible and is generally found in studies of human capital accumulation (Rebelo et al., 1998, Heckman et al., 1998). On the other hand, the role of good investment in human capital generation is independent of schooling as well as firm effects. The reason is that good investment is a sort of worker's private investment such as investment in laptop, health etc. Therefore, the firm effects should not matter here. It is also made independent of schooling in order to focus on the effects of schooling on time investment in training.

The assumption that training T_{it} and investment of goods I_{it} are additively separable in human capital production is a weaker assumption as compared to the earlier assumed complementarity. Given that there are no clear evidence on whether goods and time investment are complementary or not in the human capital production, it is apt to start with a more neutral formulation. This type of formulation is also important for analysis. For example, in times of no or less technical innovations, one would expect individuals to spend more time working rather than getting training to update their skills. But with technical changes taking place, then one is expected to rely more on training to learn and increase her skills. Similarly, in such a formulation one can check that if T_{it} and I_{it} are not direct complements in the human capital production, whether worker's optimal behavior can make them to move in one direction with schooling. Such a situation is usually called strategic complementarity in the literature. Moreover, the assumption that goods investment also matters in human capital accumulation is more common in the literature and can be supported through several empirical evidences. For example, a numbers of studies and health related reports show that income and different physical and mental diseases have a negative relation (Braveman et al., 2010, Kanervisto et al., 2011, NCHS, 2012, Pollack et al., 2013, Woolf et al., 2015, Lubetkin and Jia, 2017). This implies that high income people are investing in their health besides consumption or that consumption gives better health also in addition of its utility effects. Better physical and mental health must mean high work efficiency for this group of people.

Finally, regarding spill-over effects, studies usually assume that part of the human capital accumulation results from peer effects, i.e. average human capital stock in the economy affects individual human capital accumulation (Munandar, 2008, Aloi and Tournemaine, 2013). But here, a slight deviation is taken to enrich this by instead introducing firm effects. The reasons for this are two fold. First, many empirical works show that part of workers' productivity and wages are explained by firm's characteristics[10], which they call firm effects. Second, workers mostly interact or are conscious about others' performance and learn from each other in

[10]The firm size effects on productivity and wages are well established. For example, Brown and Medoff (1989) study shows persistent firm size effects on wages which remains even after one control for workers' selection and firm's characteristics. Davis and Haltiwanger (1991), using data on more than three million US plants between the 1963-86 period, show that observable plant characteristics account more successfully for the inter-industry wage differential than the observable workers characteristics. According to Oi and Idson (1999), workers in larger firms are more productive. Card et al. (2013), using a matched employer-employee panel data set from the west Germany, find that increasing wage inequality in west Germany over the last 25 years is approximately equally explained by increased heterogeneity between workers, increasing heterogeneity between establishments, and increased assortative matches between the two. Similarly Barth et al. (2016), using several data sets, find that more than 70% of the increased variance in earnings among the US individuals between the 1970s-2010 is associated with increased variance of average earnings among the establishments where they work. Similarly, firm characteristics like its technological complexity, capital-labour ratio etc. are found to have effects on wages for given worker's characteristics.

their organisation. The amount of learning from other workers in an organization not only depends on the skills of workers but also on the overall environment of the organization. For example, some organizations may provide more opportunities to interact and share ideas whereas other may not do so. Similarly, the technology status and other characteristics of the firm should matter[11]. Thus, the best way to introduce spill-over effects may be to say that it comes from the production function of the firm or economy. The production function is a much broader term and can incorporate the status of both human capital and technology of the economy.

Let μ_{it} and η_{it} be the Lagrange multipliers for (3.6) and (3.7). Ignoring the non-negativity constraint, the Lagrange of consumer i problem becomes

$$\begin{aligned}\sum_{t=0}^{\infty} \phi_i^t \Bigg[\frac{[c_{it}^{\rho_i}(1-x_{it}-T_{it})^{1-\rho_i}]^{1-\sigma_i}}{1-\sigma_i} + \mu_{it}[(w_t x_{it} - T_{it})h_{it} - I_{it} - c_{it}] \\ + \eta_{it}[S_i\,(T_{it}h_{it})^{\upsilon}\left(\frac{Y_t}{N}\right)^{1-\upsilon} + \varphi I_{it} - \delta h_{it} - h_{it+1}]\Bigg].\end{aligned} \tag{3.10}$$

The first-order conditions for consumer's i problem become

$$\mu_{it} = \rho_i[c_{it}^{\rho_i} l_{it}^{1-\rho_i}]^{-\sigma_i} c_{it}^{\rho_i-1} l_{it}^{1-\rho_i}, \tag{3.11}$$

$$\mu_{it} = \varphi\eta_{it}, \tag{3.12}$$

$$(1-\rho_i)[c_{it}^{\rho_i} l_{it}^{1-\rho_i}]^{-\sigma_i} c_{it}^{\rho_i} l_{it}^{-\rho_i} = S_i \upsilon \eta_{it} T_{it}^{\upsilon-1} h_{it}^{\upsilon}\left(\frac{Y_t}{N}\right)^{1-\upsilon} - \mu_{it}h_{it}, \tag{3.13}$$

$$(1-\rho_i)[c_{it}^{\rho_i} l_{it}^{1-\rho_i}]^{-\sigma_i} c_{it}^{\rho_i} l_{it}^{-\rho_i} = w_t\mu_{it}h_{it}, \tag{3.14}$$

$$\eta_{it} = \phi_i[w_{t+1}x_{it+1} - T_{it+1}]\mu_{it+1} + \phi_i\eta_{it+1}[S_i\upsilon T_{it+1}^{\upsilon} h_{it+1}^{\upsilon-1}\left(\frac{Y_{t+1}}{N}\right)^{1-\upsilon} - \delta], \tag{3.15}$$

[11]The assumption that spill-over effects result from average human capital only thus seems to be narrow and cannot account for the technology differences etc. in different firms/countries.

and the transversility condition $\lim_{\tilde{T}\to\infty} {\phi_i}^{\tilde{T}} \eta_{i\tilde{T}} h_{i\tilde{T}+1} = 0$. Note that $\tilde{T}$ denotes a specific time period instead of training. Note from the utility function that the non-negativity constraint is not binding. Equation (3.11) gives optimality condition for consumption whereas (3.12) and (3.13) are the optimality conditions for good and time investment in human capital, respectively. Equation (3.14) determines the equilibrium labour supply of the worker. Using (3.11) and (3.12), (3.13) becomes

$$\frac{1-\rho_i}{\rho_i} c_{it} = \frac{S_i \upsilon}{\varphi} T_{it}^{\upsilon-1} l_{it} h_{it}^{\upsilon} \left(\frac{Y_t}{N}\right)^{1-\upsilon} - l_{it} h_{it}. \tag{3.16}$$

Similarly, from (3.14) one can get

$$\frac{1-\rho_i}{\rho_i} c_{it} = w_t l_{it} h_{it}. \tag{3.17}$$

Finally, putting (3.11) and (3.12) in (3.15), the Euler equation of consumer optimisation becomes

$$\left(\frac{c_{it+1}^{\rho_i} l_{it+1}^{1-\rho_i}}{c_{it}^{\rho_i} l_{it}^{1-\rho_i}}\right)^{\sigma_i} \frac{c_{it+1}^{1-\rho_i} l_{it+1}^{\rho_i-1}}{c_{it}^{1-\rho_i} l_{it}^{\rho_i-1}} = \phi_i \varphi \left[w_{t+1} x_{it+1} - T_{it+1}\right] + \phi_i \left[S_i \upsilon T_{it+1}^{\upsilon} h_{it+1}^{\upsilon-1} \left(\frac{Y_{t+1}}{N}\right)^{1-\upsilon} - \delta\right]. \tag{3.18}$$

3.3.3 Competitive Equilibrium

In this section, competitive equilibrium of the model economy is defined. Section 3.3.4 then defines the balanced growth path of the economy. A sequence

$E = (w_t, Q_t, Y_t, \{(c_{it}, T_{it}, I_{it}, x_{it}, l_{it}, h_{it+1})/i = 1, 2, 3, .., N\})_{t=0}^{\infty}$

is defined as a competitive equilibrium of the model economy if the following conditions hold

(i) For $t \geq 0$, Q_t solves the firm's profit maximisation problem (3.2) given w_t and technology,

(ii) For each $i = 1, 2, 3, .., N$, the sequence $\{c_{it}, I_{it}, T_{it}, x_{it}, l_{it}, h_{it+1}\}_{t=0}^{\infty}$ solves consumer i's utility maximisation problem (3.5-3.8) given w_t and technology,

(iii) The factor market clears in every period, i.e. $Q_t = \sum_{i=1}^{N} h_{it} x_{it}$ for all $t \geq 0$,

(iv) The good market clears in every period, i.e. $\sum_{i=1}^{N}(c_{it} + h_{it}T_{it} + I_{it} - \frac{1}{N}\Pi_t) = Y_t$ for all $t \geq 0$.

The following conditions are necessary and sufficient to ensure the above defined competitive equilibrium

$$Y_t = A_t^{1-\beta} Q_t^{\beta}, \quad Q_t = \sum_{i=1}^{N} h_{it} x_{it}, \quad c_{it} \geq 0, \quad T_{it} \geq 0, \quad x_{it} \geq 0, \quad h_{i0} > 0, \tag{3.19}$$

$$w_t = \beta A_t^{1-\beta} Q_t^{\beta-1}, \tag{3.20}$$

$$\begin{aligned}
\left(\frac{c_{it+1}^{\rho_i} l_{it+1}^{1-\rho_i}}{c_{it}^{\rho_i} l_{it}^{1-\rho_i}}\right)^{\sigma_i} \frac{c_{it+1}^{1-\rho_i} l_{it+1}^{\rho_i - 1}}{c_{it}^{1-\rho_i} l_{it}^{\rho_i - 1}} &= \phi_i \varphi \left[w_{t+1} x_{it+1} - T_{it+1}\right] + \phi_i \left[S_i \upsilon T_{it+1}^{\upsilon} h_{it+1}^{\upsilon-1} \left(\frac{Y_{t+1}}{N}\right)^{1-\upsilon} - \delta\right], \\
\frac{1-\rho_i}{\rho_i} c_{it} &= \frac{S_i \upsilon}{\varphi} T_{it}^{\upsilon-1} l_{it} h_{it}^{\upsilon} \left(\frac{Y_t}{N}\right)^{1-\upsilon} - l_{it} h_{it}, \\
\frac{1-\rho_i}{\rho_i} c_{it} &= w_t l_{it} h_{it}, \\
h_{it+1} &= S_i \left(T_{it} h_{it}\right)^{\upsilon} \left(\frac{Y_t}{N}\right)^{1-\upsilon} + \varphi I_{it} - \delta h_{it}, \\
l_{it} + T_{it} + x_{it} &= 1, \\
c_{it} + I_{it} &= w_t x_{it} h_{it} - T_{it} h_{it} + \frac{\Pi_t}{N}.
\end{aligned} \tag{3.21}$$

Note that summing the individual budget constraint in (3.21) over the N individuals one gets $\sum_{i=1}^{N} c_{it} + \sum_{i=1}^{N} h_{it} T_{it} + \sum_{i=1}^{N} I_{it} - \Pi_t = w_t Q_t$. Putting wage and profit from (3.3) and (3.4) and cancelling common terms gives $C_t + \sum_{i=1}^{N} h_{it} T_{it} + I_t = Y_t$. This means that total consumption, direct training costs and good investment in human capital in the economy is equal to total output.

3.3.4 Balanced Growth Path Solution

The balanced growth path of the model economy is defined as one where Q_t, Y_t, c_{it}, I_{it} and h_{it} grow at equal and constant rate g, and the time variables T_{it}, l_{it} and x_{it} grow at zero rate[12]. The balanced growth is given by $g = g_A$ for $\beta \in (0,1)$. That is, at the balanced growth path, all the unbounded variables grow at the growth rate of the exogenous technology[13]. Now, for the balanced growth solution, define transformed variables as $\hat{Y}_t = \frac{Y_t}{A_t}$, $\hat{Q}_t = \frac{Q_t}{A_t}$, $\hat{c}_{it} = \frac{c_{it}}{A_t}$, $\hat{h}_{it} = \frac{h_{it}}{A_t}$ and $\hat{I}_{it} = \frac{I_{it}}{A_t}$. Also define two ratios as $z_{it} = \frac{\hat{c}_{it}}{\hat{h}_{it}}$ and $m_{it} = \frac{\hat{I}_{it}}{\hat{h}_{it}}$. Using this and the condition of equal growth, all the equations can be transformed in a similar fashion. For example, the per unit wage Equation (3.3) can be written as $w_t = \beta \hat{Q}_t^{\beta-1}$, which is time invariant at the balanced growth path. Similarly, dividing the maximised profit in (3.4) by A_t one can get $\hat{\Pi}_t^* = \frac{\Pi_t^*}{A_t} = (1-\beta)\hat{Q}_t^{\beta}$. A balanced growth equilibrium will involve a set of stationary values: $\left\{\hat{c}_i^*, \hat{h}_i^*, \hat{I}_i^*, x_i^*, l_i^*, T_i^*\right\}_{i=1}^{N}$, w^*, $\hat{Y}^*$ and $\hat{\Pi}^*$. Using this, one can transform the system in (3.21) to the following stationary system of equations for each i at the balanced growth path equilibrium

$$
\begin{aligned}
z_i^* + m_i^* &= w^* x_i^* - T_i^* + \frac{\hat{\Pi}^*}{N\hat{h}_i^*}, \\
(1+g)^{\kappa_i} &= \phi_i \varphi \left[w^* x_i^* - T_i^*\right] + \phi_i \left[S_i \upsilon (T_i^*)^{\upsilon} (\hat{h}_i^*)^{\upsilon-1} \left(\frac{\hat{Y}^*}{N}\right)^{1-\upsilon} - \delta\right], \\
\frac{1-\rho_i}{\rho_i} z_i^* &= \frac{S_i \upsilon}{\varphi} (T_i^*)^{\upsilon-1} (\hat{h}_i^*)^{\upsilon-1} \left(\frac{\hat{Y}^*}{N}\right)^{1-\upsilon} l_i^* - l_i^*, \\
\frac{1-\rho_i}{\rho_i} z_i^* &= w^* l_i^*, \\
1+g &= S_i (T_i^*)^{\upsilon} (\hat{h}_i^*)^{\upsilon-1} \left(\frac{\hat{Y}^*}{N}\right)^{1-\upsilon} + \varphi m_i^* - \delta, \\
l_i^* + T_i^* + x_i^* &= 1,
\end{aligned}
\tag{3.22}
$$

[12]Note that the firm effects in the human capital production ensure equal balanced growth. If the level of any individual's human capital forges ahead of the average, its growth rate slows down and convergence of human capital growth rates occurs.

[13]To obtain this, divide output $Y_{t+1} = A_{t+1}^{1-\beta} Q_{t+1}^{\beta}$ by Y_t to get $(1+g) = \frac{A_{t+1}^{1-\beta} Q_{t+1}^{\beta}}{A_t^{1-\beta} Q_t^{\beta}}$. Rearranging this, one can get $(\frac{A_{t+1}}{A_t})^{1-\beta} = (1+g)(\frac{Q_t}{Q_{t+1}})^{\beta}$, which is equal to $(1+g_A)^{1-\beta} = (1+g)^{1-\beta}$. This gives $g_A = g$.

where $\kappa_i = \rho_i\sigma_i - \rho_i + 1$, $z_i^* = \frac{\hat{c}_i^*}{\hat{h}_i^*}$ and $m_i^* = \frac{\hat{I}_i^*}{\hat{h}_i^*}$. In (3.22), there are six equations in six unknowns. This system is solved by first expressing all the variables as functions of the per unit wage rate w^*. The system in (3.22) can be expressed as given below

$$
\begin{aligned}
T_i^* &= \left(\frac{(w^*+1)\varphi}{S_i\upsilon}\right)^{\frac{1}{\upsilon-1}} \frac{\hat{Y}^*}{N\hat{h}_i^*},\\
x_i^* &= \frac{(1+g)^{\kappa_i} + \delta\phi_i + \varphi\left(\frac{(w^*+1)\varphi}{S_i\upsilon}\right)^{\frac{1}{\upsilon-1}} \phi_i \frac{\hat{Y}^*}{N\hat{h}_i^*} - S_i\upsilon\left(\frac{(w^*+1)\varphi}{S_i\upsilon}\right)^{\frac{\upsilon}{\upsilon-1}} \phi_i \frac{\hat{Y}^*}{N\hat{h}_i^*}}{w^*\varphi\phi_i},\\
m_i^* &= \frac{(1+g) + \delta - S_i\left(\frac{(w^*+1)\varphi}{S_i\upsilon}\right)^{\frac{\upsilon}{\upsilon-1}} \frac{\hat{Y}^*}{N\hat{h}_i^*}}{\varphi},\\
z_i^* &= \frac{\rho_i\left[w^*\phi_i\varphi - (1+g)^{\kappa_i} - \delta\phi_i\right]}{(1-\rho_i)\varphi\phi_i},\\
\hat{h}_i^* &= \frac{\hat{Y}^*\left[(1-\beta)(1-\rho_i)\varphi\phi_i + (1-\rho_i)(1-\upsilon)S_i\left(\frac{(w^*+1)\varphi}{S_i\upsilon}\right)^{\frac{\upsilon}{\upsilon-1}}\phi_i\right]}{N\left[w^*\rho_i\phi_i\varphi + (1+g)(1-\rho_i)\phi_i - \rho_i\delta\phi_i - (1+g)^{\kappa_i}\right]},\\
l_i^* &= 1 - T_i^* - x_i^*,
\end{aligned}
\tag{3.23}
$$

where $\frac{\hat{Y}^*}{N\hat{h}_i^*}$ can be obtained by rewriting the penultimate equation in (3.23) and is a function of w^* only. Also equilibrium output is $\hat{Y}^* = \left(\frac{\beta}{w^*}\right)^{\frac{\beta}{1-\beta}} = (\sum_{i=1}^N x_i^*\hat{h}_i^*)^\beta$. From (3.23) one can get comparative statics like $d\hat{(.)}_i^*/dS_i$, $d\hat{(.)}_i^*/d\phi_i$, $d\hat{(.)}_i^*/d\sigma_i$ for all variables under a given value of $w^* > 0$. But before that, the following paragraphs elaborate on how to solve the above system and to ensure that a unique balanced growth path equilibrium can exist for all $w^* > 0$.

To this end, first note that wage rate is $w^* = \beta(\hat{Q}^*)^{\beta-1} = \beta(\sum_{i=1}^N x_i^*\hat{h}_i^*)^{\beta-1}$. The wage rate can be re-written as $w^{*\frac{1}{\beta-1}} = \beta^{\frac{1}{\beta-1}}(\sum_{i=1}^N x_i^*\hat{h}_i^*)$. From $\hat{h}_i^*$ value in (3.23) one can see that $\hat{Y}^*$ can be taken as common from the sum $\sum_{i=1}^N \hat{h}_i^* x_i^*$. Thus wage rate can be written as $w^{*\frac{1}{\beta-1}} = \hat{Y}^*\beta^{\frac{1}{\beta-1}}(\sum_{i=1}^N x_i^*\hat{\hat{h}}_i^*)$. Note that $\hat{\hat{h}}_i^*$ is equal to $\hat{h}_i^*$ with the exception that $\hat{Y}^*$ in $\hat{h}_i^*$ is replaced with one in $\hat{\hat{h}}_i^*$ because $\hat{Y}^*$ is taken out of the sum as common. Now, multiply both sides of $w^{*\frac{1}{\beta-1}} = \hat{Y}^*\beta^{\frac{1}{\beta-1}}(\sum_{i=1}^N x_i^*\hat{\hat{h}}_i^*)$ by

$\frac{\beta}{\hat{Q}^*}$ and using definitions of $\hat{Y}^*$ and wage rate, one can get

$$\beta = w^* \left(\sum_{i=1}^{N} x_i^* \hat{\bar{h}}_i^* \right) \tag{3.24}$$

After putting values from (3.23), this equation becomes a function of a single variable; the per unit wage rate w^*. Thus, Equation (3.24) can be solved for w^*. Once w^* is determined, one can put back into (3.23) to get balanced growth competitive solution values of $\hat{h}_i^*$, $\hat{I}_i^*$, $\hat{c}_i^*$, T_i^*, x_i^*, l_i^* and the individual level wage distribution. Before solution, it is shown that the system in (3.23) has a unique solution in the following proposition

Proposition 7. *Given that $g, \sigma_i > 0$ and $\upsilon, \rho_i, \varphi, \delta, \varphi \in (0,1)$, the system in (3.23) can have only one unique solution for all $w^* > 0$.*

Proof. To prove this, note from $\hat{h}_i^*$ value in (3.23) that the numerator of $\hat{\bar{h}}_i^*$ is positive for all $w^* > -1$ irrespective of S_i. Moreover, the denominator of $\hat{h}_i^*$ or $\hat{\bar{h}}_i^*$ is negative for small values of w^* since $(1+g)\phi_i < (1+g)^{\kappa_i}$. Now, suppose $\hat{w}^* > 0$ is the value of wage rate at which the denominator of $\hat{h}_i^*$ or $\hat{\bar{h}}_i^*$ in (3.23) just turns from negative to positive. The cases of $w^* \geq \hat{w}^*$ and $w^* < \hat{w}^*$ are considered in two steps. For this, x_i^* and $\hat{\bar{h}}_i^*$ values are first substituted into (3.24) to get

$$\beta = \sum_{i=1}^{N} \frac{(1+g)^{\kappa_i} \hat{\bar{h}}_i^* + \delta \phi_i \hat{\bar{h}}_i^* + \varphi \left(\frac{(w^*+1)\varphi}{S_i \upsilon} \right)^{\frac{1}{\upsilon - 1}} \frac{\phi_i}{N} - S_i \upsilon \left(\frac{(w^*+1)\varphi}{S_i \upsilon} \right)^{\frac{\upsilon}{\upsilon - 1}} \frac{\phi_i}{N}}{\varphi \phi_i}. \tag{3.25}$$

(i) The case of $w^* \geq \hat{w}^*$: At $\hat{w}^*$ value, $\hat{\bar{h}}_i^*$ must be very large, i.e. reaches infinity when wage is such that $\hat{\bar{h}}_i^*$ denominator approaches zero from the right hand side, as the numerator is positive and far greater than the denominator at this point. Moreover, $\hat{\bar{h}}_i^*$ is continuous in $w^* \geq \hat{w}^*$ and approaches zero as w^* goes to infinity[14]. Additionally, the last two terms in (3.25)'s numerator give negative value. To see this, impose $\varphi \left(\frac{(w^*+1)\varphi}{S_i \upsilon} \right)^{\frac{1}{\upsilon - 1}} = S_i \upsilon \left(\frac{(w^*+1)\varphi}{S_i \upsilon} \right)^{\frac{\upsilon}{\upsilon - 1}}$ on the last two terms to get $1 = w^* + 1$. This implies that $\varphi \left(\frac{(w^*+1)\varphi}{S_i \upsilon} \right)^{\frac{1}{\upsilon - 1}} < S_i \upsilon \left(\frac{(w^*+1)\varphi}{S_i \upsilon} \right)^{\frac{\upsilon}{\upsilon - 1}}$ for all $w^* > 0$.

[14]The denominator of $\hat{\bar{h}}_i^*$ goes to infinity and numerator approaches the constant first term.

So, the right hand side of (3.25) is very large at $\hat{w}^*$ whereas the left hand side is a constant of less than one. Also, (3.25) is continuous for all $w^* \geq \hat{w}^*$. Then as w^* goes to infinity, the right hand side of (3.25) approaches a negative value. This ensures one crossing point at a positive wage rate. Also, note that $\hat{h}_i^*$ is positive for such wage values. But then the positive wage implies that x_i^* must also be positive for all $w^* \geq \hat{w}^*$.

(ii) The case of $0 < w^* < \hat{w}^*$: Now, it is shown that no equilibrium can exist for $0 < w^* < \hat{w}^*$. From (i), $\hat{h}_i^*$ must be negative for $0 < w^* < \hat{w}^*$ for all i. Also, a negative $\hat{h}_i^*$ implies negative $\frac{\hat{Y}^*}{N\hat{h}_i^*}$ from the second last equation in (3.23), as positive w^* must mean $\hat{Y}^* > 0$ from the definitions of w^* and $\hat{Y}^*$ above. Now, from (3.23), I can show that $x_i^* > 0$ holds even if wage is such that $\hat{h}_i^*$ is negative. To see this, note that $\varphi\left(\frac{(w^*+1)\varphi}{S_i\upsilon}\right)^{\frac{1}{\upsilon-1}} < S_i\upsilon\left(\frac{(w^*+1)\varphi}{S_i\upsilon}\right)^{\frac{\upsilon}{\upsilon-1}}$ for all $w^* > 0$. But then the negative $\frac{\hat{Y}^*}{N\hat{h}_i^*}$ in the x_i^* value implies that $x_i^* > 0$ must hold for all $0 < w^* < \hat{w}^*$. But $x_i^* > 0$ and $\hat{h}_i^* < 0$ for all i implies that w^* cannot be positive, so a contradiction again. □

3.3.5 Comparative Static Analysis

This section analyses the effects of pre-job schooling differences on consumption, human capital accumulation, wages and time allocation between training, work and leisure. As mentioned earlier, schooling S_i can have effects on the patience level of individuals $\phi(S_i)$, can affect preference between consumption and leisure $\rho(S_i)$, it can affect inter-temporal elasticity of substitution $\sigma(S_i)$, and can enter the human capital production function directly as in the above specification. All these possibilities will be considered in the following analysis.

Proposition 8. *Given a solution with $w^* \geq \hat{w}^*$, then*

(i) $\hat{h}_i^*(w^*)$, $\hat{I}_i^*(w^*)$, $x_i^*(w^*)$ and per hour gross wage $w^*\hat{h}_i(w^*)$ are decreasing in ϕ_i and $1/\sigma_i$ on the balanced growth path,

(ii) $T_i^*(w^*)$ and $l_i^*(w^*)$ are increasing in ϕ_i and $1/\sigma_i$ on the balanced growth path.

Proof. See the appendix. □

Since, we expect education to increase the patience level of individuals or to make them sacrificing their current for better future career, ϕ_i must be a positive function of schooling S_i if pre-job schooling has any effects on the discount factor. Similarly, one expects schooling to increase the inter-temporal elasticity of substitution $1/\sigma_i$ (Havranek et al., 2015). Given this, the result about T_i^* in Proposition 8 is consistent with the evidence that pre-job schooling increases training incidence during job. Thus, if ϕ_i and $1/\sigma_i$ are increasing in pre-job schooling S_i, then the model predicts that pre-job schooling can lead to increasing time investment in training through its effects on individuals' preferences, ceteris paribus. However, then the rest of the results in Proposition 8 are not plausible at all. If pre-job schooling S_i results in an increasing ϕ_i and $1/\sigma_i$ alone, then Proposition 8 implies that gross total wage $w^*\hat{h}_i(w^*)x_i^*(w^*)$ and per hour gross wage $w^*\hat{h}_i(w^*)$ must fall in pre-job schooling. But there are abundant evidence that show that wages increase in pre-job schooling. The numerical solution below plots the US earning distribution by education categories which clearly shows that wages strictly increase in education. Similarly, the result that human capital falls in schooling is difficult to support.

Similar result emerges if one considers the effects of pre-job schooling on preferences about leisure versus consumption. For example, from (3.23), one can see that $\frac{\partial \hat{h}_i^*(w^*)}{\partial \rho_i} > 0$ implies $\frac{\partial T_i^*(w^*)}{\partial \rho_i} < 0$ and vice versa. This implies that per hour gross wage $w^*\hat{h}_i(w^*)$ must be negatively associated with training T_i^* and schooling S_i, which is not true as discussed above. Similarly, $\frac{\partial \hat{h}_i^*(w^*)}{\partial \rho_i} > 0$ implies $\frac{\partial \hat{I}_i^*(w^*)}{\partial \rho_i} > 0$ from (3.23). Thus, neither human capital and time investment in training nor goods and time investment in human capital move in the same direction under ρ_i effects of pre-job schooling. As discussed later, this result holds even if one keeps the direct costs of training at zero in the individual's budget constraint. For more detailed results on the effects of differences in preferences over leisure versus consumption, see Figures B.15-B.17 in the appendix. Next, the study moves to the case of schooling having a direct role in human capital accumulation to get the following results:

Proposition 9. *Given a solution with $w^* \geq \hat{w}^*$, then*

(i) $l_i^*(w^*)$ is independent of S_i on the balanced growth path,

(ii) $\hat{h}_i^*(w^*)$, $\hat{c}_i^*(w^*)$, $T_i^*(w^*)$ and per hour gross wage $w^*\hat{h}_i(w^*)$ are increasing while $x_i^*(w^*)$ is decreasing in S_i on the balanced growth path.

Proof. See the appendix. □

These results seem more plausible in terms of the observed relation between training, consumption, gross wages and pre-job schooling. It clearly shows that consumption, human capital and gross wages strictly increase in schooling. Another interesting result from this analysis is the finding that leisure does not increase in schooling even though work hours are falling. But before detail discussion, the study first proceeds to a complete numerical solution of the system in (3.23). Solution of (3.23) for a given distribution of schooling S_i and parameter values can give complete distributions of $\hat{I}_i^*$, $\hat{h}_i^*$, $\hat{c}_i^*$, the per unit wage rate $w^* = \beta(\hat{Q}^*)^{\beta-1}$ and individual's net of training cost wage distribution $w^*\hat{h}_i^*[x_i^* - T_i^*]$.

3.3.6 Numerical Solution

For numerical solution, the study first focuses only on the direct role of schooling in human capital accumulation. Thus, it is necessary to specify the values of S_i, ϕ, σ, β, ρ, φ, υ, δ and growth rate g_A. Regarding the choice of schooling distribution, formal schooling is usually distributed among three to five categories depending on the country's educational structures[15]. In this study, a comparison between the model predicted wage distribution and actual wage distribution of the US workers

[15] Altonji and Spletzer (1991) study distributes it into vocational, less than two years college, more than two years college, college and advance degree. Green (1993) categorises schooling into vocational, O-level, A-level, higher degree and other qualifications. Brunello (2001) study, based on ECHP data, distributes education into three stages; lower secondary, upper secondary and higher.

will be carried. The US wage data by education categories is usually reported for five categories of schooling; less than a high school diploma, high school graduates with no college, some college or associate degree, bachelor's degree only and advanced degree. Keeping this, five discrete values of schooling are taken with uniform distribution as $S_i = \{0.5, 0.6, 0.7, 0.8, 0.9\}$. The choice of uniform distribution will allow to see clearly whether the variables increase at an increasing or decreasing rates in schooling. The values of less than one for schooling is to make it consistent with the human capital theory, where the schooling efficiency parameter is chosen between zero and one (Lucas, 1988, Rebelo, 1991).

The value of β is set equal to 0.7 and growth rate is set equal to 2.2%. This approximates the share of labour and annual average growth of US economy's real per-capita GDP over the last five decades. The value of discount factor ϕ is set equal to 0.97 (King and Rebelo, 1999). Regarding the value of σ, there is a lot of variation in the results of different studies, giving σ between and including 1.0 and 2.0 (King and Rebelo, 1999, Havranek et al., 2015). The meta analysis of Havranek et al. (2015) shows that, based on $1,429$ estimates, the mean inter-temporal elasticity of substitution of US economy is 0.594. Keeping this in view, the value of σ is set at 1.6. Regarding depreciation of human capital, empirical studies of job training show that the effects of training on wages last from four to thirteen years (Lengermann, 2000, Lillard and Tan, 1986). If this is used as a signal of skill depreciation, then it means that the human capital depreciation can range from 8.0 to 25.0 percent. This study keeps δ equal to 0.2 implying 20.0 percent depreciation[16].

The remaining parameter values are taken arbitrarily and then changed to check that the results hold for a wide range of values. In the literature, the spill-over effects' share ranges between 0.2 to 0.8 (Abowd et al., 1999, Aloi and Tournemaine, 2013, Card et al., 2013, Barth et al., 2016). Keeping this, for the baseline model the value of υ is set equal to 0.5 and then increased to 0.8 later on. With regard to φ value, it is initially set at 0.6 and then changed to see its impact on choices

[16]Note that parameter values can have magnitude effects but it does not have effects on the distribution of variables which is the interest of this study.

of individuals. Similarly, the value of ρ is set to 0.5 and subsequently changed to see its effects. Although, different schooling level will take different years of time, it is assumed that each individual has same amount of time to allocate between work, training and leisure. This enables one directly compare the differences in training due to schooling differences. To accommodate for differences in schooling time, one can rewrite the human capital accumulation equation as $h_{it+1} = e_i(1 - S_i)\left(T_{it}h_{it}\right)^{v}\left(\frac{Y_t}{N}\right)^{1-v} + \varphi I_{it} - \delta h_{it}$. Where e_i now shows the efficiency of schooling. But this will not change the following results when schooling has enough efficiency effects. Also, for the sake of simplicity, single individual is assumed in each schooling category so that N is equal to five in the model economy. Finally, in the first part of the analysis, it is assumed that schooling directly enters human capital production, as in (3.7), but has no effects on ϕ, σ, ρ etc. Such a specification is called human capital efficiency effects of schooling.

Results with human capital efficiency effects of schooling

The results for the above baseline parameter values are reported in Figure 3.1. More detailed results based on actual values are shown in Figures B.3-B.5 in Appendix B.5. All the variables, except work time, are normalized in Figure 3.1 in order to plot all the curves in a single figure. From Figure 3.1, it can be seen that time in job training, consumption and human capital are increasing in schooling, while good investment and work time decrease in pre-job schooling[17]. Time in training, consumption and human capital increase linearly in schooling and training curve is steeper than the consumption and human capital functions. Furthermore, the results on time allocation are consistent with the results shown in Proposition 9 above. Also, optimal training and leisure times are far less than the work time. The

[17]Later on, we will see that good investment can increase in schooling when the firm effects are weaker. The plausibility of the results of decreasing goods investment in human capital as a function of schooling can depend on the type of training investment and nature of organization. If a person got theoretical knowledge of certain technology during schooling and the firm has that machine and provides its practical training, then this may kills the worker incentives to buy her own machine. On the other hand, if goods investment is considered as investment in health for example, then it is more likely that the highly qualified will invest more in health.

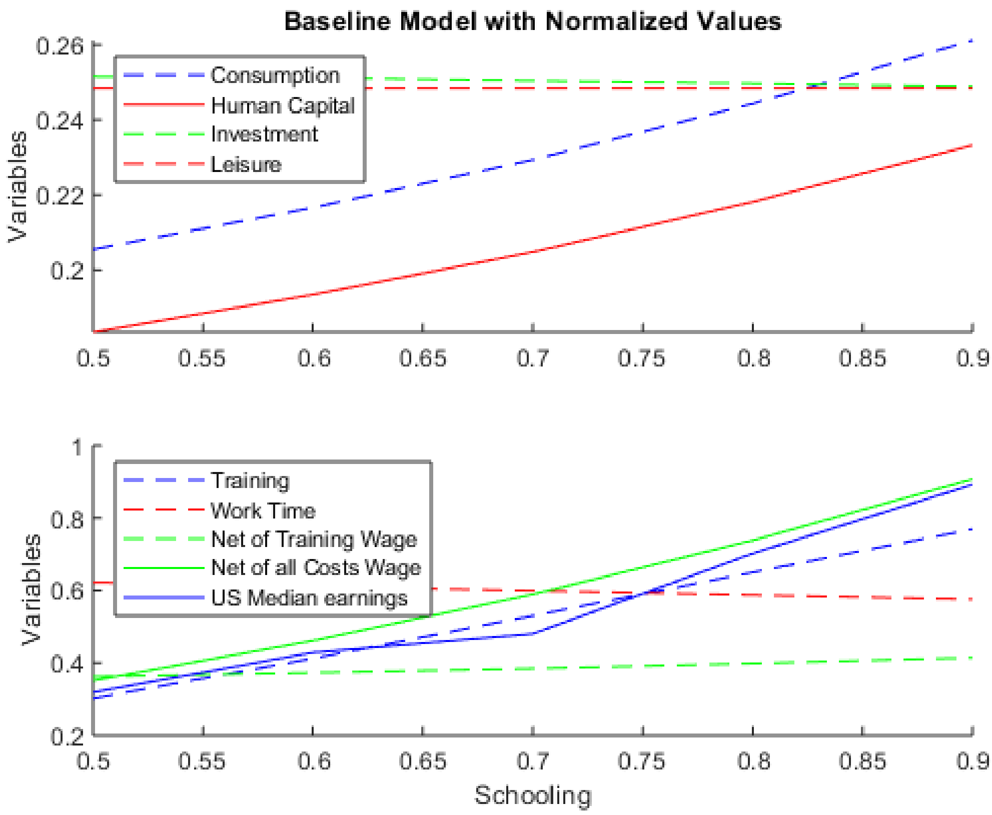

Figure 3.1: Distributions of Variables under the Baseline Parameter Values

important feature of the results is that the distributions of consumption, investment and human capital are relatively flat as compared to the setting with exogenous labour supply as shown in Appendix B.3. For example, both consumption and human capital shares of the individual with highest schooling are close to 22.6% whereas for the individual with second highest schooling it is 21.1%. For the lowest schooling individual the shares are close to 17.8%. In total training, the share of the highest schooling individual is 28.9%. At the lowest schooling category, the share is 11.3% and at the second highest degree level the share becomes 24.4% approximately. This difference is moderate if one compares it with the exogenous labour supply case in the appendix.

While work time decreases in schooling, leisure is independent of schooling. Thus, although highly qualified individuals spend relatively less time in work, this does not mean that they spend relatively more time in leisure. Rather they spare this time for learning during job period. This is important new prediction of the model.

Unlike the commonly held belief that high income (here highly qualified) people spend more time in leisure, the model predicts that they actually spend this time in learning. This result is quite interesting and needs empirical investigation. In the real world, it is difficult to find a highly qualified individual who spends relatively more time in leisure on average if leisure is considered as a complete waste of time or activities like cooking for fun, stay at home, wasteful time with friends etc. If one takes the example of a qualified academic who may stay for less hours in office. The academic in no way wastes as much time as usually a low qualified worker do on average. What this means is that this individuals spend more time in learning, though it may be in an informal ways like reading at home etc., as compared to low schooling workers. Thus, the finding of decreasing labour supply may not be straight away interpreted as implying more time in leisure for the highly qualified individuals.

Overall, the model with endogenous labour supply correctly predicts the empirical evidence presented in Section 3.2, that is, highly qualified individuals spend more time in training. It also correctly predicts that highly qualified (high income) people spend less time in work. But unlike the popular belief, the model predicts that the optimal behavior for the highly qualified would be to spend this extra time in training; making leisure independent of skill level of individuals. If a truly realistic specification, this implies that models which split worker's time between leisure and work only, can lead to misleading predictions about time allocation and wage distribution. If the training time is treated as leisure, it will lead to under estimation of wages for the highly qualified individuals. Similarly, treatment of training as work time will lead to misleading wage function as training time has different effects on take home wages. Moreover, the model with baseline parameters predicts the life cycle of workers much better than their daily life. For example, if the post-schooling life is normalized as one, the model predicts that approximately 60.0% of the time is spent in work, 5.0% in job training and 35.0% in leisure (which will henceforth be treated as retirement life).

Regarding wage, two individual level wage distributions, net of training cost $(w^*x_i^* - T_i^*)\hat{h}_i^*$ and net of all costs $(w^*x_i^* - T_i^*)\hat{h}_i^* - I_i^*$, are drawn in Figure 3.1 (see appendix for more details). One can see from the figure that both net of training cost and net of training plus good investment costs wage functions are linearly increasing in schooling. Thus, net wage function is increasing in schooling even if per unit wage w is decreasing in aggregate effective human capital. Although increasing, the net wage function is not as steep as it is in the case of exogenous labour supply in the appendix. The reason is that with endogenous labour supply, the low schooling individuals partly compensate for their low skills by working for more time as compared to highly qualified individuals. The curve with kinks in Figure 3.1 plots the median weekly earnings of US full time wage and salary workers by education categories, reported by the US Bureau of Labor Statistics in its 4th quarterly, 2017. To plot the curves together, the US median weekly earnings are normalised. When compared to the US weekly earnings by education in Figure 3.1 or in the appendix, one can see that the model generated net of all costs wage distribution can approximately match the data on average. However, the net of training cost wage function is flatter and cannot concisely explain the earning inequalities among different education categories in the US. But it is not clear entirely if the US earning distribution is reporting gross earning or making any corrections for training and development expenditures of firms or workers. Additionally, although not drawn in the figures, the model generated gross per hour wage $w^*\hat{h}_i^*$ also closely matches the US earning distribution. Finally, in the specification it is assumed that workers pay all costs of training. If workers pay only part of the training cost, then $(w^*x_i^* - T_i^*)\hat{h}_i^*$ must become steeper than the one drawn.

As mentioned above, $\hat{I}_i^*$ is decreasing in schooling which many can disagree with, say for example high schooling individuals spend more on health, laptops etc. To this end, the firm effects are weaken to show that investment $\hat{I}_i^*$ can also increase in schooling, along with all the other results above still intact. This is true despite the fact that I_i and T_i are additively separable in the human capital production. Firms

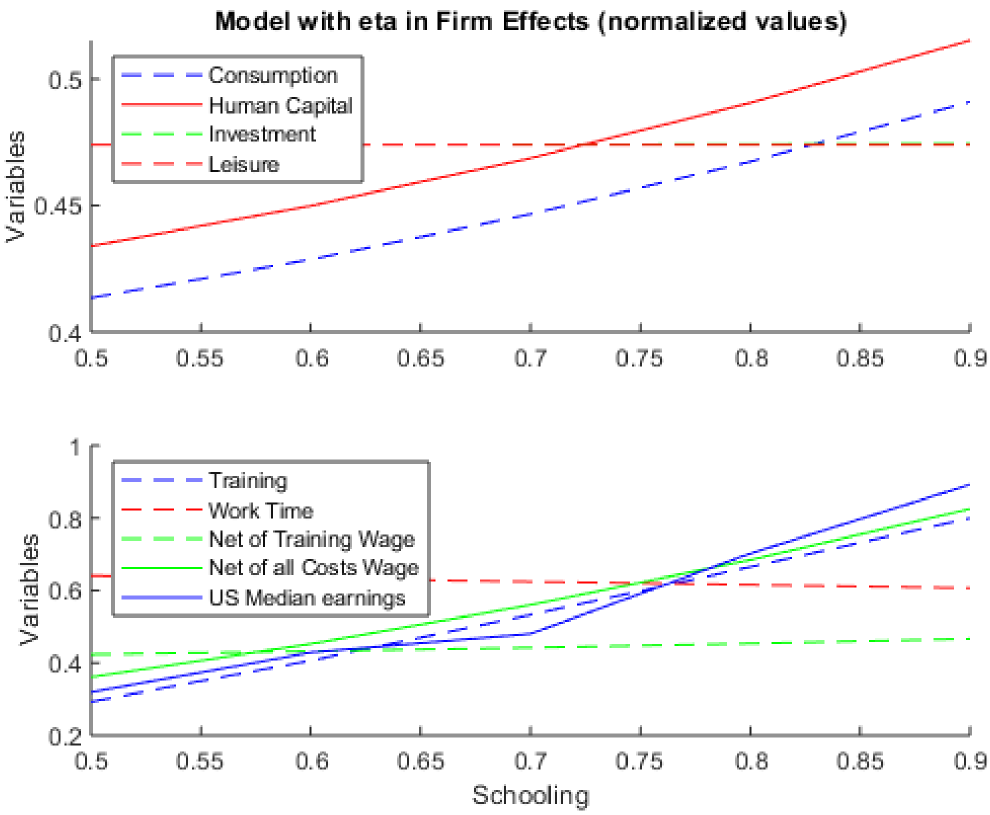

Figure 3.2: Distributions of Variables with a Change in the Firm Effects

effect were introduced earlier as $(\frac{Y_t}{N})^{1-\upsilon}$. Now $\eta \in (0,1)$ is introduced into it to make it weaker, so that it now looks like $(\frac{\eta Y_t}{N})^{1-\upsilon}$. The only change in the equations system (3.23) is that $\frac{\hat{Y}^*}{N\hat{h}_i}$ is replaced by $\frac{\eta\hat{Y}^*}{N\hat{h}_i}$. With such a specification, as the value of η decreases the investment starts getting flatter and for η value of 0.8 it becomes a positive function of schooling. For the results that will be reported, the value of η is set equal to 0.6. Moreover, the value of δ is also changed from 0.2 to 0.05. Note that the results hold with any δ value but δ of 0.2 was the highest possible in the literature. Rest of the parameters are the same as earlier. The results are reported in Figure 3.2. More detailed results based on actual values are shown in Figures B.6-B.8 in the appendix.

As is clear from the figures, now goods investment also increases in schooling and all the other earlier results still hold. The reason is that weaker firm effects make investment of time in job training relatively less attractive, as a given amount of time in training now leads to relatively less increase in future human capital.

On the other hand, the highly skilled are needed to pay relatively more for a given amount of training as is clear from the consumer budget constraint. Keeping this, the highly educated workers shift some of the resources to goods investment and their investment now increases in schooling. To better understand this, one can keep δ equal to 0.2 like earlier and can compare the results with the baseline results above. In such an exercise, one will see that now the individuals invest relatively less time in training and relatively more goods investment as compare to the baseline results. In the baseline results, the firm effects were so strong that it paid to the qualified worker to invest more in training, since firm effects complement training and the worker get free lunch. In doing so the worker was not feeling the need for goods investment in future human capital accumulation because it is not without cost. Regarding net wages, it still strongly increase in pre-job schooling and with δ value of 0.05 the net of all costs wage can explain almost 80% of variation in the US workers' weekly earnings by education categories. Thus, in the empirically realistic case of increasing T_i^* and $\hat{I}_i^*$ in schooling, pre-job schooling differences can explain a significant variation in wages by schooling categories given that schooling enters the human capital accumulation equation directly. But it cannot explain all the variation in wages, signalling that other factors play role in explaining workers earning gaps.

Before going to discuss the effects of a change in υ, the effects of a change in the values of φ and ρ are briefly discussed here. For this analysis, all the other parameters take the baseline values. When φ increases, all the baseline results hold. But now individuals choose more goods investment in human capital and decreases job training time. This is plausible and highlights the fact that a relative change in the effectiveness of different sources of human capital accumulation can have important effects on the choices of time and goods investments in human capital. Thus, one can see that important changes might have occurred due to the rapid technical changes over the last few decades. For example, if training importance is increasing with technical innovations, then workers may reduce their goods investment and go for more training. Moreover, when ρ value decreases from 0.5 to 0.25, the model predicts an average work time of 0.37. This approximates the daily working time of around 8.0 hours. This is an expected result as low value of ρ means more utility from leisure which induces workers to spend less time in work. With such a decrease in ρ, workers also spend less time in training; average of 0.025 compared to 0.053 in the baseline model[18]. Another important change that happens with a decrease in ρ value is that now I_i is increasing in pre-job schooling. The reason is that the training time creates more dis-utility for the individuals with low ρ, so they shift to goods investment. Also, average work time under $\rho = 0.5$ was 0.6, which now decreases to 0.37. This makes the firm effects $(\frac{Y_t}{N})^{1-\upsilon}$ weaker and force the highly qualified to spend relatively less on T_i^* and allocate more resources to $\hat{I}_i^*$. Remember that, although the highly qualified spend relatively less on T_i^*, but T_i^* still increases in pre-job schooling.

Next, the value of υ is increased, which implies high power of effective training time and corresponding weaker share of firm effects in the human capital accumulation. In this experiment, the value of υ is increased from 0.5 to 0.8 which is the highest value in the literature. The results for this experiment are shown in Figures B.9-B.11 in the appendix. As is clear from the figures, now consumption and human capital distributions are flatter as compared to the baseline distributions. The individuals now invest relatively more goods and spend less time in training as compared to the baseline results. Low training time at $\upsilon = 0.8$ than at $\upsilon = 0.5$ is because the positive externalities in the form of firm effects are weak now and the worker is forced to pay for most part of the human capital accumulation. Although low in absolute value, time in training function is now more steep function of schooling; one can see this if the two curves are drawn on separate figures. Thus, the increase in υ pushes the qualified individuals to aggressively invest more time

[18]Many would think that ρ value of 0.25, which predicts daily work life, may be small compared to empirical evidence. To this end, the alternative way for getting this result with high ρ value can be to introduce a parameter, as was done in the firm effects above, into the utility function. For example, one may write the utility function as $\frac{[c_{it}^{\rho}(ol_i)^{1-\rho}]^{1-\sigma}}{1-\sigma}$ with $o \in (0, 1)$. Note that the weighting parameter is a common practice in additive utility function.

in training as compared to low qualified individuals. But the decrease in the power of firm effects hurts the individuals enough to force them invest overall less time in training. Moreover, the negative sloped working hours function is now flatter and everyone is working for more hours as compared to the baseline results.

Net of job training wage $(w^* x_i^* - T_i^*)\hat{h}_i^*$ now decreases in schooling because training is a highly steep function of schooling and the high skilled are needed to pay relatively more for a given training time. But the overall net wage function is increasing because high skilled individuals are investing less goods in human capital. Moreover, in addition to $\upsilon = 0.8$ if the value of $\eta = 0.6$ is kept, the goods investment is still a negative function of schooling. Thus, this experiment shows that with $\eta = 0.6$, there is a critical value of υ between 0.5 and 0.8 such that goods investment function changes its sign with respect to schooling level. Furthermore, in what ever way the firm effects may be specified, the above comparative statics imply that one can get a variety of wage distribution, ranging from linear to strongly convex, by changing the relative share of firm effects in the human capital accumulation. With most reasonable parameter values, the generated gross per hour wage function $w\hat{h}_i$ can explain a substantial part of the US workers' weekly earning distribution by education categories. Next, the study checks the effects of changing discount factor as were explored in Proposition 8.

Results with patience effects of schooling

To check the effects of changing patience level, in this subsection the study assumes that $S_i = S = 1$ in (3.7) but the time discounting parameter ϕ is a function of schooling S_i. It takes the values of ϕ as $\phi(S_i) = \{0.93, 0.94, 0.95, 0.96, 0.97\}$. All the other parameters take the baseline values. The results for this exercise are shown in Figures B.12-B.14 in Appendix B.9. As is clear from the figures, now consumption, human capital, goods investment, work time, gross and net wages strictly decrease in $\phi(S_i)$. Why consumption and human capital decrease in patience level of the individuals? It is because to cut down the cost of training the highly patience individuals choose less human capital and instead choose high leisure. By doing so, they are supposed to pay less for a given amount of training, as training costs are deducted according to $T_i\hat{h}_i$. On the other hand, the increasing leisure time makes them overall better off as compared to the impatience individuals; the discounted utility of patience individual is more than the utility of the impatience individuals. The results of varying σ instead of ϕ are exactly the same as are shown in Proposition 8. Similar results also emerge from introducing differences in individual's preference over labour-leisure choices as are shown in Figures B.15-B.17.

Note that from the above specification but with exogenous labour supply, it can be shown that balanced growth consumption, human capital, investment and net wage is increasing in schooling if schooling makes the agents more patience or changes their inter-temporal elasticity of substitution, even if schooling has no direct effects on human capital production. Thus, the above results are in contradiction with the results for ϕ changes with exogenous labour supply. In the case of endogenous labour supply, the highly educated individuals can simply choose high leisure or less working hours so that all or part of the schooling effects on wages can be offset. Introducing physical capital can weaken the ϕ effects in a model with exogenous labour supply specification but cannot offset it totally, so that the positive ϕ and human capital investment relation can still hold with exogenous labour supply. This implies that in certain cases, the assumption of exogenous labour supply may not be a simplifying one. Rather, it can have important implications for the results. Overall, this exercise implies that within the given specification, to get realistic consumption, human capital, investment and wage distribution, pre-job schooling must have direct effects on human capital accumulation.

3.3.7 Discussion

These results give important insights into the effects of differences in schooling on individual's behavior towards time in training, leisure, working time, human capital and consumption. Firstly, to get the empirically proven positive school-job training relation, it does not matter whether individual schooling affects individual's preferences or their ability of human capital accumulation. But to get plausible schooling and consumption, human capital and wage relation, pre-job schooling must affect individual's ability of human capital accumulation. Under the assumption that schooling improves human capital production efficiency during job, the results suggest that training time can increase in schooling even if training affects individual's utility negatively and high schooling individuals pay relatively more for a given amount of training time. Thus, the condition that schooling improves human capital production efficiency is both necessary and sufficient to generate the observed positive relation between pre-job qualifications and job training, consumption and wages.

Is it plausible to assume that pre-job schooling facilitates the individual in human capital accumulation during the job? Some evidence can be presented to support this assumption. For example, Lengermann (2000) categorises individuals from NLSY data on the basis of scores on the Armed Services Vocational Aptitude Battery (ASVAB). ASVAB measures verbal, mathematical and mechanical ability as well as computational speed. He shows that average educational attainment increases with respondents' test scores. Mean years of schooling for the low, middle, and high scorers ranges as 11.7, 12.9, and 14.5 respectively. This means that highly educated are usually high ability individuals, and thus they must be more efficient in on-the-job learning and skills accumulation. Heckman et al. (1998) present further evidence to support the notion that pre-job schooling should have efficiency effects in human capital accumulation.

Secondly, under the assumption that pre-job schooling has human capital efficiency effects, the model predicts that highly educated individuals spend more time in training and they spare this time from working for relatively less time. Thus, the optimal behavior for each individual is to choose equal time for leisure. This implies that if we see high skilled people working for relatively less time, we cannot conclude that they necessarily spend more time in leisure. This result is quite interesting and needs empirical investigation. Today, many household surveys explicitly ask questions about time in work and in leisure. Some surveys, like British Household Panel Survey, even ask questions about the utilization of leisure time as well. This data can be used to see the relation between individuals' qualifications and leisure time. Another implication of this prediction is that one should try to split time into work, learning and leisure instead of labour-leisure only in models where one is interested to study inequality. As, against the prediction of typical labour-leisure models, in real world, it is difficult to find a highly qualified individual who is spending relatively more time in leisure on average if leisure is considered as a complete waste of time.

Additionally, this study implies that the introduction and strength of firm effects can be used to match possibly many observed wage distributions, and the corresponding parameter values can be estimated from the model. Moreover, changes in the relative effectiveness of training time versus goods investment in human capital have important re-allocation effects. These results are important keeping in view the changing nature of jobs over the last few decades. For example, in jobs where more physical energy is needed, people may spend more on consumption to remain healthy instead of getting training. Similarly, when production techniques are not changing, then workers may not feel the need of more training. But the technological changes over the last few decades have greatly increased the importance of mental abilities in job performance, and consequently, the importance of physical work has decreased. This implies that the workers today may like more training as compared to thirty years ago for their given schooling. But then, given the positive association between schooling and job training, this also means that wage level and growth gaps between different schooling categories should also be high today as compared to thirty years ago. This is what a number of studies have consistently reported (Autor, 2014).

Finally, what would happen to the results if training has no direct costs? In such a case, the worker does not need to pay $h_{it}T_{it}$ any more as costs of training. In terms of the solution in equations system (3.23), the term $w + 1$ will be replaced by w everywhere. One can see from this that the introduction of training cost has important magnitude effects on training, work hours and human capital accumulation. Training time now increases more aggressively in pre-job schooling. Particularly, then $w^* x_i^* \hat{h}_i^*$ becomes the net wage of the worker. In such a case, the resultant wage function can become even steeper with the baseline parameter values. On the other hand, if direct cost of training exists but the firm is forced to pay part of it, then the results will lie within the two extreme cases, i.e. no direct training cost versus training cost and workers are bearing all of it. But in both these cases, the result that leisure is independent of pre-job schooling is not going to change.

3.4 Conclusion

A number of micro data-based studies have confirmed that workers who enter jobs with relatively high pre-job schooling also get more training during jobs. The positive association between schooling and training incidence implies that pre-job skill heterogeneities amplify during one's job. This can increase wage inequality in the post-entry period of job depending on the quality and quantity of training that an individual undertakes and on the costs of such investment. Keeping these facts, the current study tries to answer why and under what conditions the more educated will choose more investment of time and goods in human capital, and then its implications for wage changes. Following the existing human capital accumulation literature, the study considers both goods and time investment in human capital during the job. But unlike earlier studies, in this study the worker needs to pay the direct costs of training besides the foregone wage income when the worker is in training.

The results show that an individual training time increases in pre-job schooling even if utility is decreasing in training and high skilled individuals pay relatively more for a given training time, given that pre-job schooling improves the efficiency of training time in future human capital generation. Moreover, the positive school-training relation can hold when the pre-job schooling has preferences effects; in the form of affecting patience level of the worker as a consumer or affecting preferences about labour-leisure choices. But under the preference effects of schooling, the resultant distributions of human capital and wages among individuals with different schooling level are not plausible.

When pre-job schooling has direct role in human capital accumulation, then working time decreases but leisure time is independent of individual's schooling level. In other words, the model predicts that highly qualified individuals spend the time they get from working less hours in learning instead of leisure. This prediction is new, seems more realistic and highlights the need for more detailed study of time allocation by workers with different qualifications and abilities. In real world, if we account for on-the-job and off-the-job learning and time spent in other healthy activities, we can easily see that highly educated individuals spend at least as much time in these activities as the poor low-skilled workers do on average. If one splits time between work and leisure only then such a model's prediction of decreasing work hours will, by definition, mean more leisure time and can lead to misleading results in certain cases. Under the assumption that pre-job schooling has direct role in human capital accumulation, the model generated wage distribution is increasing in schooling and can explain a substantial portion of US workers' median earning distribution by education categories.

To conclude, this study shows that the assumption that pre-job schooling improves the worker's efficiency in human capital production during job is enough not only to establish positive school and time-in-training association but also generates increasing human capital, consumption and wage functions in pre-job schooling on the balanced growth path. This assumption can generate positive school-training association under the extreme condition that the worker is paying all direct and indirect costs of training. The resultant wage distribution can explain major part of US workers earning distribution by education categories. The model also predicts that all individuals, irrespective of their human capital, optimally spend equal time in leisure. This result is in contrast to the prediction that high income individuals should spend more time in leisure but seems to predict the real world well. Finally, the result that when the good investment becomes less productive, individuals increasingly rely on training for human capital generation is important. The technological changes over the last few decades have greatly increased the importance of mental abilities in job performance, and consequently the importance of physical work is decreasing. This implies that the workers today may like more training as compared to thirty years ago for their given schooling and income. But this must mean that the wage level and growth gaps between different schooling categories should also be high today as compared to thirty years ago. This is what a number of empirical studies is consistently showing.

Chapter 4

Wage Return of Training with Heterogeneity Robust Methods

4.1 Introduction

There is a vast literature trying to estimate the wage return to job related training. Many studies during the 1990s and the 2000s conducted for the US, UK and European countries find substantial return to job training[1]. However, studies during the 2010s that are based on data sets created through random experiments estimate insignificant wage return from training. Moreover, studies during the 1990s and the 2000s mostly apply fixed-effects approaches on multi-periods panel data sets. However, many recent working papers show that the standard fixed-effects methods in multi-periods data setting can be misleading if treatment effects, training effects in this case, are heterogeneous across units and overtime.

To analyse these concerns, this study re-estimates the effects of training on wages by applying some of the newly developed heterogeneity robust estimation techniques on data sets from British Households Panel Surveys. For comparison purpose, the study first carries estimations with the standard two-way fixed-effects method that is not robust to heterogeneous treatment effects. The results from the application

[1] Detail Discussion on this and possible econometric issues with these studies is given in Section 4.2.

of standard fixed-effects method show that training incidence has significant positive effects on wages. Employer provided training seems to have strong wage effects as compared to government training schemes or university degrees and diplomas. Although the effects of training are positive and significant from the fixed-effects method, it is not as high as reported in some previous studies. A given incidence of training increases the real or nominal gross wage in the range of 0.38% to 1.03%. The current study controls for many possible confounders, and in many cases its square and cubic roots, like job tenure, qualifications, job satisfaction, age, unemployment spells, dependent children, job status, managerial duties, firm size, non-labour income etc. Controlling for these confounders decreases the coefficient of training incidence in many cases, but the magnitude of change is not really big. Unlike time-varying confounders, the inclusion of fixed-effects and individual-specific time trends changes the coefficient of training incidence and the overall R-squared by a significant amount.

Among the heterogeneity robust estimation methods, this study applies three approaches[2]; namely those developed by Abraham and Sun (2018), Callaway and Sant'Anna (2018) and Imai et al. (2018). The estimations from all three approaches suggest that training has positive and significant effects on workers' wages. The average wage return to training from these estimations is higher than the wage return estimates from standard fixed-effects methods. For example, the average wage effects of a single period training on the trained ranges in 2.9-3.8% when the Callaway and Sant'Anna (2018) method is applied on the developed panels. According to the Abraham and Sun (2018) and Imai et al. (2018) methods, the wage return of training is even greater. For instance, the wage return from Imai et al. (2018) approach ranges between 6% to 9% for training incidence(s) that happens in a single year over the sample period. More importantly, the effects of training pick-up over time in the post-training period. All three methods suggest that the positive wage effects of training persist for many years, and that the long run effects

[2]Detail discussion of these approaches is given in Sections 4.2 and 4.3.

of training are higher than the contemporaneous or immediate post-training period effects. The overall picture this exercise gives is that training is likely to have economically and statistically significant positive effects on workers' wages. If there is any bias in the standard fixed-effects, the bias seems to be downward, giving low return estimates of training[3]. Another message from the study is that one needs to study the wage effects of training over mutli-periods, rather considering only the immediate post-training period for analysis.

After the introduction, the next section reviews the literature on training return estimates and discusses the possible econometric and data issues. Section 4.3 briefly highlights the newly developed heterogeneity robust methods. Section 4.4 presents data and results from the standard two-way fixed-effects method for the overall panels. Section 4.5 presents results of the newly suggested heterogeneity robust estimation methods. Section 4.6 concludes the study.

4.2 Literature Review

There is a vast empirical literature that estimates return of training to firms and workers since the 1990s. This research started with the emergence of data on direct training measures, which itself resulted due to the dissatisfaction with the use of indirect measures such as tenure and experience to measure human capital formation. Many earlier studies conducted for the US, UK and European countries found substantial return to training. The return estimates to one week of training is in many cases as high as the return to one year of formal schooling (Blundell et al., 1996, Loewenstein and Spletzer, 1999, Kuckulenz and Zwick, 2003, Schøne, 2004, Frazis and Loewenstein, 2005). For example, Frazis and Loewenstein (2005) use the National Longitudinal Survey of Youth (NLSY) dataset and estimate different forms of regressions for the 1979-2000 period. They find a minimum return of 40%

[3]The bias of the two-way fixed-effects is much clear by looking at Abraham and Sun (2018) application to BHPS panel in Table 4.4. There, one can see that the standard dynamic two-way fixed-effects gives negative return to training as compared to the positive return under Abraham and Sun (2018) application.

to 50% to one week of private training. Lengermann (2000) uses the NLSY panel for the 1979-1993 period and shows substantial earning effects of long spelled company training. The estimated return is 4.4% in the first year of training and increases to 8.2% after 9 years.

Using fixed-effects estimation and data from Follow-Up to the School Leavers Survey (FSLS), Parent (2003) shows that for men, participation in employer supported training increases wage return by 10.34% in Canada. For UK, Blundell et al. (1999) use National Child Development Survey (NCDS) data for the 1981-1991 period and report 5% return to an incidence of employer provided training under the fixed-effects and 6.5% return with the instrumental variable approach. Fixed-effects estimates from an employer provided training course that does not lead to a qualification is 12%. Haelermans and Borghans (2012) provide a meta-analysis of 38 studies and show that the average wage return to training is 2.6%. This type of studies were used to argue that there might be under-investment in training due to credit constraints of workers, and firms' hesitation to invest in trainings out of fear of worker mobility.

However, many later studies that use data from random experiments find no significant wage return to training (Leuven and Oosterbeek, 2004, 2008, Görlitz, 2011, Schwerdt et al., 2012, Hidalgo et al., 2014, Görlitz and Tamm, 2016). Leuven and Oosterbeek (2004), exploit Dutch tax system to use regression discontinuity design for estimation of wage return to training. This tax provision shares 40% of training cost for only those employees who are forty years and above aged. They find that the coefficient of training is small and insignificant in OLS estimations, and under 2SLS method its sign is not clear besides insignificance. The other papers compare the wages of trained to those untrained who wanted to participate but had to cancel training due to random reasons, or randomised experiments data. Leuven and Oosterbeek (2008) got such sample through two survey questions where respondents wanted to participate in training but could not because of random events such as illness, family circumstances etc. They find that the coefficient of training drops to very low level and insignificant when they narrow down the comparison group to such individuals. Schwerdt et al. (2012) use data from randomized field experiment of voucher training in Switzerland and find no impact of voucher program on earnings, employment and further education. However, firm financed training has a positive impact in this study.

In another similar study, Hidalgo et al. (2014) use a randomized experiment of CINOP Centre of Expertise in which a training voucher of €1000 was given to low-skilled workers in the Netherlands. They could not find any significant wage or mobility effects of such training. The coefficient of training is positive but insignificant in OLS estimation when they include controls. When they use treatment status as an instrumental variable, the coefficient of training becomes negative and insignificant. Görlitz and Tamm (2016) compare the wage and employment outcomes of training recipients financed by a German training voucher program to those who got the voucher but did not get training due to random reasons. Their two-way fixed-effects show negative but insignificant impact of training on wages. Such findings are not limited to studies using random experiments only. Many studies, using household panel surveys, also found either low or insignificant wage return to private training. For a later example, Albert et al. (2010) find insignificant return estimate to training in six European countries using fixed-effects estimations.

This approach has a definite advantage as here the data is either from random experiments or close-to-random experiment. However, in most of these studies, the comparison group and sample size is small (Leuven and Oosterbeek, 2008, Görlitz, 2011). For example, sample size in Leuven and Oosterbeek (2004) ranges from 149 to 275 for the narrow comparison groups. In Görlitz (2011), control group of those who wanted to participate in training but could not has a size of 150 in an overall sample of 1603. In Leuven and Oosterbeek (2008), the completely random control group size is 77 only whereas in Görlitz and Tamm (2016) they are 164 individuals in a total sample of 938. Such a small sample size raises doubt on the results about significance level, particularly in regard to training return. Unlike formal schooling, training programs can vary substantially from firm to firm, in terms of whether the training is about general or firm-specific skills, whether it has led to a qualification or not and whether the worker is working under long term contracts or can leave the incumbent firm easily. All these aspects are likely to lead to substantial variations in wage return to training. Thus, to get reliable results about the significance level of training return a relatively large sample would be desirable.

Another weak point of these latter studies is the data consideration for only one of post-training period. Most of these studies use the wage data in the immediate post-training period. For example, in Leuven and Oosterbeek (2004, 2008) and Schwerdt et al. (2012), data are collected within 12 months of the training. In Hidalgo et al. (2014), the voucher is given in 2006 and data are collected in 2007 and 2008. In Görlitz and Tamm (2016), post-training period data is collected within a maximum of 14 months after the training. However, private training is likely to affect wages in the long term. In most cases, the training firm pays the direct costs of training but is likely to recover training cost by not passing the training benefits to workers in the immediate post-training period. For example, although training has no effects on wages in Görlitz and Tamm (2016), the training participants are more often engaged in non-routine analytic tasks in post-training period. This is likely to increase their future wages. Similarly, Lengermann (2000) study shows that wage return from training is higher in later years as compared to its immediate return.

Despite all these weaknesses of the latter studies, there is one more concern which needs to be addressed before one can believe on the high return estimates from studies during the 1990s and the 2000s. Most of these studies use fixed-effects regressions[4], in some cases instrumental variables methods, to estimate training

[4]One reason behind the routine use of fixed-effects estimation is that it is highly improbable to find an instrument that is correlated with training but orthogonal to the error term of the regression to control for selectivity bias (Leuven and Oosterbeek, 2008). Another reason for the use of fixed-effects methods is the common belief that it is equivalent to Difference-in-Difference (DID) method (Bertrand et al., 2004, Angrist and Pischke, 2009). If it is equivalent to DID and if the parallel trends assumption of DID holds, then fixed-effects method is more general in the sense that it allows to control for unobservable individual characteristics. Given that it controls for both observable and unobservable individual characteristics, it is assumed that the common

return (Blundell et al., 1999, Parent, 1999, 2003, Arulampalam and Booth, 2001, Pischke, 2001, Schøne, 2004, Frazis and Loewenstein, 2005). However, recent econometric studies[5] highlight that the standard two-way fixed-effects (with both unit and time fixed-effects) methods may give misleading results of treatment effects even if the common trends assumption of DID is met. The reason is that the treatment coefficient in the two-way fixed-effects model is a weighted sum of units (groups) and time average treatment effects[6]. These weights depend on the estimation method and the weights can be negative if treatment effects are heterogeneous (Borusyak and Jaravel, 2017, Abraham and Sun, 2018, de Chaisemartin and D'Haultfœuille, 2018, Goodman-Bacon, 2018, Imai et al., 2018). Goodman-Bacon (2018) shows that negative weights arise when treatment effects vary over time and bias the regression estimates away from the true treatment effects. He cautions against the use of two-way fixed-effects in multi-periods setting for summarising treatment effects into a single aggregate parameter when treatment effects vary over time.

Regarding the implication of such heterogeneity, Gibbons et al. (2014) replicate eight influential papers published in American Economic Review between 2004-2009 and show that the true average treatment effects are different from fixed-effects estimation in five papers at 5% significance level. Treatment effects heterogeneity is found in seven of the papers. Thus, this study shows that treatment heterogeneity is prevalent and leads to misleading results. Goodman-Bacon (2018) uses his proposed decomposition strategy to replicate Stevenson and Wolfers' (2006) two-way fixed-effects analysis of the effects of unilateral divorce reforms on female suicide in thirty seven US states. The fixed-effects method gives treatment effects of -3.8, whereas his proposed decomposition suggests that the true effect is -5.0. In another study, de Chaisemartin and D'Haultfœuille (2018) develop heterogeneity robust estimator

trends assumption of DID must meet, so that the resultant coefficient on the treatment variable must present the true causal effects of treatment in fixed-effects methods.

[5]Most of these studies, such as Gibbons et al. (2014), Borusyak and Jaravel (2017), Abraham and Sun (2018), Athey and Imbens (2018), Callaway and Sant'Anna (2018), de Chaisemartin and D'Haultfœuille (2018), Goodman-Bacon (2018), Imai et al. (2018) and Imai and Kim (2019b), are in working paper stage from different US schools.

[6]See Equation (4.2) in Section 4.3.

and apply it to see the effects of NTV channel on the share of voting in favour of opposition parties in Russia, and the effects of newspapers on voter turnover in presidential elections of US. They show that their estimand gives opposite sign of treatment effects as compared to the standard fixed-effects method in one of the application and in the second application gives same sign as fixed-effects but significantly different in magnitude. Their findings hold for both fixed-effects and first-difference regressions.

The finding that fixed-effects method may assign negative weights to some of the units or time average treatment effects is particularly important in terms of training. In training, one possibility is that the employer might make the workers pay for training by accepting relatively low wages during or in the immediate post-training period. In such a case, the true contemporaneous effects of training may be negative. Then, the possible negative weights can turn such negative effects into a misleading positive coefficient on the training variable in fixed-effects estimation[7]. Moreover, Imai and Kim (2019b) identify many (so for understated) assumptions of fixed-effects methods that must hold for it to estimate the true causal treatment effects. The two important assumptions are that past treatment does not affect current outcomes and past outcomes do not affect current treatment. In the case of training, it is highly possible that training has dynamic effects, violating the former assumption. One way this could be adjusted is including certain lags of the treatment (training) variable. However, Abraham and Sun (2018) show that if treatment effects are heterogeneous, then including lags does not return causally interpretable estimands. The estimand associated with a particular lag may represent a non-convex average of treatment effects on the treated from all periods and weights for certain periods may be negative as well. Abraham and Sun (2018) argue that the non-convex average and the possibility of negative weights are enough to result in treatment effects estimate which has opposite sign compared to the true treatment

[7]On the other hand, if the true unit/group-wise return is positive but fixed-effects method assign negative weights due to heterogeneity, then the fixed-effects coefficient of training will under-estimate the true wage return of training.

effects[8]. The latter assumption is more problematic in training as is well known from the so-called Ashenfelter's dip; which refers to the observation that employees earnings drop dramatically before they participated in training. This implies that past outcomes can affect participation in training. Violation of this assumption violates strict exogeneity, which is necessary for causal identification in fixed-effects estimations (Imai et al., 2018, Imai and Kim, 2019b).

Given this discussion, this study re-estimates the effects of training on wages by applying some of the newly developed estimation techniques on data sets from the British Household Survey. More specifically, the methods developed by Callaway and Sant'Anna (2018), Abraham and Sun (2018) and Imai et al. (2018) along with the standard two-way fixed-effects are used to analyse the effects of training on wages in the United Kingdom. Callaway and Sant'Anna (2018) propose a multi-periods weighted DID estimator where the weights are inverse probabilities of selection into treatment conditional on pre-treatment observed covariates as in Abadie (2005). In addition of relaxing the parametric restrictions of fixed-effects, the benefit of Callaway and Sant'Anna (2018) is that it allows for treatment timing variation. Additionally, it weakens the common trends assumption to conditional common trends. Particularly, unlike many non-parametric techniques that give average treatment effects for each year without a single aggregate parameter, Callaway and Sant'Anna's (2018) method allows group-time treatment effects to be aggregated in many meaningful ways as will get clear later. It also allows for formal testing of the validity of pre-treatment conditional common trends assumption. Because of variation in treatment timing, this method allows to check common trends for some groups at a time where other groups are treated, and one can see the result of treatment on their wages (see Figures 4.2 and 4.3).

However, Callaway and Sant'Anna's (2018) approach allows for conditioning on only single period pre-treatment covariates. If past outcomes affect current treatment, which has been shown in Ashenfelter's dip in the case of training, then it is

[8]In their replication of Dobkin et al. (2018) analysis of hospitalization, they show that the fixed-effects estimations give opposite sign compared to their suggested approach.

desirable to control for pre-training wages (outcomes) history. Imai et al. (2018) propose a matching method that controls for past outcomes besides past treatment and other covariates histories. It first constructs a refined matched set of control units for each treated unit based on past treatment, outcomes, and covariates histories. The refinement can be performed using any standard matching method. Using the matched set, they develop weighted DID to estimate the causal effects of treatment. Thus, this method assumes common trends conditional on past treatment history, past outcomes and observed time-varying other covariates. The weak point of this method is that it requires relatively many periods' data to allow for proper lags and leads to control for distance past history. Many data observations are likely to lose during the refinement process. Though Imai et al. (2018) approach is very strong and addresses many of the concerns raised earlier, like any other matching method, it cannot control for time invariant unobservable unit effects. On the other hand, one strong point of fixed-effects regression is its ability to control for units specific unobservables. To address such concerns, this study also carries estimation through Abraham and Sun (2018). Abraham and Sun (2018) refine two-way fixed-effects to make it robust to heterogeneous treatment effects in multi-periods panel estimations.

4.3 Estimation Methods Highlights

The commonly used approach to estimate training return in multi-periods panel data is the following two-way fixed-effects regression:

$$lnW_{it} = \alpha_i + \theta_i t + \gamma T_{it} + X_{it}\beta + \epsilon_{it}, \tag{4.1}$$

where lnW_{it} is natural log of individual's gross monthly wage, α_i is time invariant individual effects, $\theta_i t$ captures individual specific time trends, X_{it} are time-varying confounders like qualifications, job tenure, job satisfaction, marital status, firm size, non-labour income etc. As in many training studies (Pischke, 2001), T_{it} denotes the level of training an individual i has received until time t. This is because training leads to permanent increase in human capital of individual i. The term θ_i captures heterogeneity in individuals wage growth and thus is expected to purge out the effects of any correlation between unobserved determinants of wage growth and training.

If the assumption of common trends holds[9] and the true training/treatment effects are constant across units and overtime and equal to say $\diamond$, then the γ in the above regression is equal to $\diamond$ (Athey and Imbens, 2018, de Chaisemartin and D'Haultfœuille, 2018). However, if treatment effects vary over time and/or across individuals (or groups), then γ can be written as (de Chaisemartin and D'Haultfœuille, 2018, Imai and Kim, 2019a)

$$\gamma = \sum_{i=1} \sum_{t=1} \omega_{i,t} \diamond_{i,t}, \tag{4.2}$$

where $\omega_{i,t}$ are weights and $\diamond_{i,t}$ are the true treatment effects on individual (group) i in time period t. Note from (4.2), that for γ to equal the true treatment effects $\diamond_{i,t}$, the weights must be properly assigned. Many of the recent working papers show that these weights are not guaranteed to equal the share of each group in the data, and even worse many of the $\omega_{i,t}$ can take negative values for reasons mentioned earlier (Borusyak and Jaravel, 2017, Abraham and Sun, 2018, Athey and Imbens, 2018, de Chaisemartin and D'Haultfœuille, 2018, Goodman-Bacon, 2018). If this is the case, Equation (4.2) clearly shows that the negative weights can change the true treatment effects into a completely misleading results. Thus, in the case of such heterogeneity of treatment effects, the common trends assumption alone is not enough to ensure estimation of the true causal effects of treatment. Imai and Kim (2019a) further show that the estimation of Equation (4.1) is not allowing for any dynamic relation between the treatment and outcomes. This would be a very strong assumption in training keeping in view the Ashenfelter's dip.

[9] This is the basic assumption of DID method and means that the average outcomes of treated and untreated should grow at a common rate in the absence of treatment.

4.3.1 Callaway and Sant'Anna (2018) Approach

Callaway and Sant'Anna (2018) estimate the average treatment effects for individuals first treated in period e at time period t

$$ATT(e,t) = E(W_t(1) - W_t(0)|G_e = 1), \tag{4.3}$$

where $W_t(1)$ is the potential outcome (wage) at time t if treated, $W_t(0)$ is the potential outcome at time t if not treated and $G_e = 1$ is a binary variable denoting that the individual belongs to the group first treated in period e (the first event period). Besides overlap and irreversibility of treatment assumptions[10], the parallel trends assumption they impose is the same as in Abadie (2005). That is, for all $t = 2, ..., T$, $e = 2, ..., T$ such that $e \leq t$

$$E(W_t(0) - W_{t-1}(0)|X, G_e = 1) = E(W_t(0) - W_{t-1}(0)|X, C = 1), \tag{4.4}$$

where X is a vector of pre-treatment covariates and C is a binary variable equal to one if an individual is in the control group. This assumption states that the average outcomes for the group first treated in period e and for the control group would have followed parallel trends in the absence of treatment conditional on covariates X. This conditional common trends assumption must hold for each group and for all times after the treatment period. Under these assumptions, the following simple weighted average recovers the group-time average treatment effects

$$ATT(e,t) = E\left[\left(\frac{G_e}{E(G_e)} - \frac{\frac{p_e(X)C}{1-p_e(X)}}{E[\frac{p_e(X)C}{1-p_e(X)}]}\right)(W_t - W_{e-1})\right], \tag{4.5}$$

where the generalized propensity score $p_e(X) = P(G_e = 1|X, G_e + C = 1)$ denotes the probability that an individual gets treated conditional on X and being a member of group G or C. The intuition behind (4.5) is this; it weights up observation from

[10]Note that in the training literature, the standard assumption is that training leads to permanent increase in human capital (Pischke, 2001). This assumption implies that the irreversibility of treatment assumption is satisfied in case of training. However, the permanent human capital formation assumption is relaxed in later estimations for robustness check.

the control group C which has similar characteristics as those found in the treatment group, and weights down observation from the control group having characteristics not being found in the treated group. Estimation of (4.5) involves two steps. In the first step, one estimates the generalized propensity score $p_e(x) = P(G_e = 1|X = x, G_e + C = 1)$ and gets the fitted values for each group, by using either Logit or Probit model. In the second step, one puts these sample estimates $\hat{p}_e(x)$ into the sample analogue of (4.5) to get $\widehat{ATT}(e,t)$.

The estimated $ATT(e,t)$ are the group-time average treatment effects for the treated. Thus, if one has panel data for say eight years, one will get twenty one group-time average treatment effects (see Figure 4.2). But in most cases, one would like to have a measure of the overall effects of a policy change. In particular, if the effects are negative for some groups (times) and positive for others, it will make it difficult to see whether the overall effects are positive or negative. The parameter γ in the fixed-effects estimation in Equation (4.1) on the other hand gives such overall effects of treatment. However, as mentioned earlier, such aggregation in fixed-effects model is based on arbitrary weighting and the weights may be negative. Callaway and Sant'Anna (2018) allow for aggregation of $ATT(e,t)$ across groups and over time in many possible ways depending on the research question and prior knowledge. The $ATT(e,t)$ can be aggregated in such a way that allows for calendar effects, selective treatment timing and possible dynamic effects of treatment (see, Section 2.3 of Callaway and Sant'Anna (2018) for details). For example, in firm-financed training, the firm may try to recover the cost of training in the immediate post-training period by not allowing the productivity gains from training to pass on to worker. In such a case, the effects of training are likely to become evident in the long run only. They suggest that in such a case one should put more weights on observations which got treatment long before time t as below

$$\theta(r) = \sum_{e=2}^{T}\sum_{t=2}^{T} 1\{t-e+1=r\} ATT(e,t) P(G=e|t-e+1=r), \tag{4.6}$$

$$\theta = \frac{1}{T-1}\sum_{r=1}^{T-1}\theta(r). \tag{4.7}$$

The first expression aggregates $ATT(e,t)$ in a way where $ATT(e,t)$ gets larger weight the greater is $t-e$. Thus, individuals who get training in early years of the sample will get more weight in the aggregation than those who are trained toward the end period in the sample. The second expression average out $\theta(r)$ to get a single summary coefficient. Though this weighting is arbitrary as in the regression models, the resulting weights are always positive and gives one more ways of aggregation and comparison. This does not affect the individual group-time average treatment effects $ATT(e,t)$ in anyway. So, if one is not happy with the weighting scheme for aggregation, the option of looking at $ATT(e,t)$ alone is always available and can give a clear picture of the treatment effects in many cases.

Finally, Callaway and Sant'Anna (2018) allow to test the pre-treatment conditional common trends. This is made possible by assuming that the conditional common trends assumption in (4.4) holds for all $t = 2, ..., T$, $e = 2, ..., T$ rather than for $e \leq t$ alone. This leads to the following null hypothesis:

$$H_0 : E(W_t - W_{t-1}|X, G_e = 1) - E(W_t - W_{t-1}|X, C = 1) = 0, \tag{4.8}$$

for all $2 \leq t < e \leq T$. Here W_t denotes the actual outcome in period t. This simply means that in the pre-treatment period, the change in group-time averages of outcomes for the treated and untreated, conditional on covariates, must not be different. That is, $ATT(e,t) = 0$ should hold for pre-treatment periods. Thus, looking at $ATT(e,t)$ in the pre-treatment periods for different groups can give enough initial idea of whether the parallel trends assumption hold in the pre-treatment periods. One can also apply a simple Wald test to check the null that the pre-training $ATT(e,t)$ is not different from zero. This is one great advantage of Callaway and Sant'Anna (2018). For example, if one has data from 2001 to 2008 and some individuals are treated in 2003 while other are treated say in 2006. Then one can see whether $ATT(e,t)$ is zero in 2002 for the group treated in 2003 and increase beyond that. At the time where this group is in the post-treatment period, the group to be treated in 2006 is still in pre-treatment time. Thus, in years 2003 to 2005, one can see how treatment is affecting the 2003 treatment group in post-treatment period and whether the $ATT(e,t)$ of the group to be treated in 2006 is still near zero. If the $ATT(e,t)$ of 2003 treated group is going above zero in 2003-2005 and $ATT(e,t)$ of group to be treated in 2006 are near zero in 2003-2005, this will increase one's faith on the effectiveness of treatment.

4.3.2 Imai et al. (2018) Approach

As mentioned earlier, in many cases past outcomes may affect current treatment. In such a situation one needs to adopt an approach which can control for differences in pre-treatment outcomes history. It is also possible that some current covariates, which are independent of current or past treatment, change differently for the treatment and control groups. In such cases, not controlling for these factors may bias the results. Imai et al. (2018) suggest an approach which has the capacity to control for both past outcomes and time-varying other covariates. This approach first constructs a matched set of control units for each treated unit based on past treatment history. Then, it refines the matched set using standard matching methods, like Mahalanobis distance, propensity score matching etc., to adjust for past outcomes and time-varying covariates. Finally, using the matched set, they apply weighted DID to estimate the causal effects of treatment. In this method, the counterfactuals are estimated as the weighted average of control units in the refined matched set. The ATT for Imai et al. (2018) is given as below

$$\begin{aligned} ATT(F,L) = E[&W_{i,t+F}(D_{i,t}=1, D_{i,t-1}=0, \{D_{i,t-l}\}_{l=2}^{L}) \\ &- W_{i,t+F}(D_{i,t}=0, D_{i,t-1}=0, \{D_{i,t-l}\}_{l=2}^{L}) | D_{i,t}=1, D_{i,t-1}=0], \end{aligned} \tag{4.9}$$

where $W_{i,t+F}(D_{i,t} = 1, D_{i,t-1} = 0, \{D_{i,t-l}\}_{l=2}^{L})$ is the potential outcome for the treated and $W_{i,t+F}(D_{i,t} = 0, D_{i,t-1} = 0, \{D_{i,t-l}\}_{l=2}^{L})$ is the potential outcome for the untreated. The term $\{D_{i,t-l}\}_{l=2}^{L}$ shows treatment history before period $t-1$. F denotes the number of leads and L denotes the number of lags. Thus, if one is interested in estimation of treatment effects for three years after the treatment administration she will have to set $F = 3$. Unlike Callaway and Sant'Anna (2018), this specification allows for both permanent treatment and future treatment reversal[11]. It is standard to expect that training leads to permanent human capital formation. But for robustness purpose, this study will present results for both possibilities, i.e. that training has human capital formation effects for just one year versus the assumption that it has human capital effects for many years. So for, to remain consistent with them, the study highlights their approach within the treatment reversal assumption scenario. The conditional common trends assumption in Imai et al. (2018) is as below

$$\begin{aligned}
&E[W_{i,t+F}(D_{i,t} = 0, D_{i,t-1} = 0, \{D_{i,t-l}\}_{l=2}^{L}) - W_{i,t-1}|D_{i,t} = 1, D_{i,t-1} = 0, \\
&\quad \{D_{i,t-l}, W_{i,t-l}\}_{l=2}^{L}, \{X_{i,t-l}\}_{l=0}^{L}] = E[W_{i,t+F}(D_{i,t} = 0, D_{i,t-1} = 0, \{D_{i,t-l}\}_{l=2}^{L}) \\
&\qquad - W_{i,t-1}|D_{i,t} = 0, D_{i,t-1} = 0, \{D_{i,t-l}, W_{i,t-l}\}_{l=2}^{L}, \{X_{i,t-l}\}_{l=0}^{L}]. \quad (4.10)
\end{aligned}$$

Thus, their conditional common trends assumption allows one to control for past treatment $\{D_{i,t-l}\}_{l=2}^{L}$, past outcomes $\{W_{i,t-l}\}_{l=2}^{L}$, and past and current values of other time-varying covariates $\{X_{i,t-l}\}_{l=0}^{L}$. After controlling for reasonable lags, one can say that this is a very weak common trends assumption. The approach proceeds as follow. First, for each treated unit at time t, it defines the matched set of control units that share the same treatment history between $t-L$ to $t-1$ periods

$$M_{it} = \{i' : i' \neq i, D_{i't} = 0, D_{i't'} = D_{it'} \quad \text{for all} \quad t' = t-1, ..., t-L\}, \qquad (4.11)$$

[11]For treatment which remains effective for F periods in the future, see Equation (19) in Imai et al. (2018).

for the treated unit i with $D_{it} = 1$ and $D_{it-1} = 0$. At this stage, each treated unit may get matched with a number of control units. For some units, the matched set may end up with no control unit. Such treated units are dropped from the analysis. The next step is to control for lagged outcomes and time-varying covariates. Imai et al. (2018) allow for using any standard matching or weighting method to perform this. For example, to refine the matched set further, one can use Mahalanobis distance as below

$$R.M_{it}(i') = \frac{1}{L}\sum_{l=1}^{L}\sqrt{(\mathbf{V}_{i,t-l} - \mathbf{V}_{i',t-l})^{\intercal}\Sigma_{i,t-l}^{-1}(\mathbf{V}_{i,t-l} - \mathbf{V}_{i',t-l})}, \tag{4.12}$$

for each matched set $i' \in M_{it}$. Here, $\mathbf{V}_{i,t'} = (W_{it'}, \mathbf{X}_{it'+1}^{\intercal})^{\intercal}$ and $\Sigma_{i,t'}$ is the sample covariance matrix of $\mathbf{V}_{i,t'}$. This uses lagged outcomes and time-varying covariates to compute standardized distance between treated unit and each control unit in the matched set and then averages it over time. Once $R.M_{it}(i')$ is obtained for each control unit in the matched set, then the matched set is refined by selecting J most similar control units that satisfy a caliper constraint. The rest of the units in the matched set get zero weight. Once the refined subset of the matched set is obtained, the counterfactual outcomes are estimated using the weighted average of the control units in the refined subset for each treated unit. Then, the following $ATT(F, L)$ are estimated:

$$\widehat{ATT}(F, L) = \frac{1}{\sum_{i=1}^{N}\sum_{t=L+1}^{T-F} Z_{it}}\sum_{i=1}^{N}\sum_{t=L+1}^{T-F} Z_{it}\left[(W_{i,t+F} - W_{i,t-1}) - \sum_{i' \in M_{it}} \omega_{it}^{i'}(W_{i',t+F} - W_{i',t-1})\right], \tag{4.13}$$

where $Z_{it} = D_{it}(1 - D_{it-1}).1\{|M_{it}| > 0\}$, and $\omega_{it}^{i'}$ denotes non-negative normalized weights with $\omega_{it}^{i'} \geq 0$ and $\sum_{i' \in M_{it}} \omega_{it}^{i'} = 1$.

4.3.3 Abraham and Sun (2018) Approach

Imai et al.'s (2018) approach is very strong and addresses many of the concerns raised earlier. For example, conditioning on past outcomes to create the matched set of control units before DID estimation should sufficiently address the possibility that the earning difference may be due to unobservable individual characteristics. This is because controlling for past wage differences is like controlling for such unobservable as well. But still, if one wants to directly control for unit-specific unobservables and at the same time takes care of the fact that the standard fixed-effects method is not robust to heterogeneous treatment effects, then one can rely on Abraham and Sun's (2018) suggested approach. They put event study design into potential outcomes framework and show that the estimand of a linear two-way fixed-effects equation like the one below can be a non-convex average of treatment effects in presence of heterogeneity:

$$W_{it} = \alpha_i + \theta_t + \sum_{r'=-T}^{T} \gamma_{r'} D_{it}^{r'} + \epsilon_{it}, \tag{4.14}$$

where r' denotes leads and lags relative to the initial treatment time as r does in Callaway and Sant'Anna (2018). Abraham and Sun's (2018) approach is based on the assumption of no reversal of treatment as is Callaway and Sant'Anna (2018). That is, treatment is an absorbing state so that it is a non-decreasing sequence of zeros and then ones. In Abraham and Sun (2018), a cohort e is defined by the event time E_i which is the time of the onset of initial treatment, i.e. $E_i = min\{t : D_{it} = 1\}$. The cohort specific ATT relative to the initial treatment time is defined as $CATT(e, r') = E[W^{e}_{i,e+r'} - W^{\infty}_{i,e+r'} | E_i = e]$, where the first term is cohort specific potential outcome r' periods relative to the initial treatment e and the second term is similar potential outcomes under no treatment scenario. To identify $CATT(e, r')$, Abraham and Sun (2018) impose the assumptions of parallel trends and no anticipatory behavior. These assumptions imply that $CATT(e, r') = 0$ for all $r' < 0$. They suggest to estimate $CATT(e, r')$ with the following interacted specification in the first step:

$$W_{it} = \alpha_i + \theta_t + \sum_{e=1}^{T-1} \sum_{r'=1-e}^{T-1-e} \delta_{e,r'} (1\{E_i = e\} D_{it}^{r'}) + \epsilon_{it}. \tag{4.15}$$

Note that this specification interacts relative time indicator with cohort indicators. The cohort which gets treatment in the first period of sample is excluded to allow for pre-treatment outcomes. Similarly, if one wants to estimate treatment effects for all those who get treatment ultimately (no units which are never treated), then one needs to exclude the last period from estimation. In the second step, their approach involves estimation of weights for each cohort as the cohort sample share in the relevant period $N_e / \sum_{e=1-r'}^{T-1-r'} N_e$. Finally, Abraham and Sun (2018) estimate what they call Interaction-Weighted (IWD) estimators by aggregating $\hat{\delta}_{e,r'}$ from step one as below

$$IWD_{r'} = \sum_{e=1}^{T-1-r'} \frac{N_e}{\sum_{e=1-r'}^{T-1-r'} N_e} \hat{\delta}_{e,r'}. \tag{4.16}$$

To complement the earlier approaches, the current study also carries estimation using this approach. The advantage of Abraham and Sun's (2018) method is that one can easily test the parallel trends and no anticipatory behaviour assumptions by using Wald test. Such testing involves the null hypothesis that $CATT(e, r') = 0$ for all pre-training periods. Following their approach, estimation is carried on the treated only. Such estimation will help avoid the doubt that there may be some basic difference in the trained and untrained which the data is not telling us, besides this method's capacity to allow for time and unit fixed-effects.

4.4 Data and Estimation Results of the Standard Fixed-Effects Method

Before estimations with the heterogeneity robust methods in Section 4.5, this section describes the data sets and also carries estimations with the standard fixed-effects method for the overall sample. This will serve two purposes. First, one can compare the results from these samples with the fixed-effects estimates from earlier studies. Second, the heterogeneity robust methods put further restrictions on the data. This study will re-estimate the fixed-effects with the restricted samples later to not only compare with the heterogeneity robust methods but also check that the results do not change significantly with changes in the samples. This section also shows how the sample size can affect the significance of training variable.

4.4.1 Data for the Analysis and Training Measures

The analysis in this paper is based on two data sets: British Household Panel Survey (BHPS) and the United Kingdom Household Longitudinal Study (UKHLS), the latter is an extension of BHPS including more observations and details. BHPS started in 1991 with 5,000 households selected at random in Great Britain. It collects detailed data on household as well as individuals in the household covering a range of social, political and economic aspects of individual and household life. In 1999, 1,500 households were added in each of Scotland and Wales. Again, it added 2,000 additional households in the year 2000 from Northern Ireland. The UKHLS started in 2009, which approximately covers 40,000 households in the United Kingdom.

The UKHLS panel used in this analysis includes waves 2 to 8, covering the years 2010 to 2016. The BHPS set includes waves 11 to 18, covering the years 2001 to 2008. The reasons for using two sets from British households are two-fold. First, a given training is not likely to affect wages for decades. Thus, a period of 6 to 8 years is reasonable time to analyse the effects of training. Also, in the literature of matching and weighting methods, it is recommended not to keep the study period very long, as longer is the time period high are the chances that other external confounding factors may affect the true causal relation. This also gives the leverage to use data sets which largely contain different individuals, thus provides the chance of cross-validation of training effects. Second, as discussed below, the information on training details are quite different in BHPS and UKHLS data. Additionally, the UKHLS sample alone is very large and thus one is safe from facing sample size issues. Given this discussion, the initial broader training question asked in BHPS waves 8 and onward is as below

(Apart from the full-time education you have already told me about) Have you taken part in any other training schemes or courses at all since September 1st year …. or completed a course of training which led to a qualification? Please include part-time college or university courses, evening classes, training provided by an employer either on or off the job, government training schemes, Open University courses, correspondence courses and work experience schemes.

Then, BHPS asks about the number of training courses and further details on the three most important trainings. It asks about training place with options; (a) at current or future workplace, (b) at former workplace, (c) employer's training centre, (d) private training centre, (e) job centre, (f) college of further or higher education, (g) adult education centre or evening institute, (h) university etc. It goes on asking about training purpose with options; (a) to help you get started in your current job, (b) to increase your skills in your current job for example by learning new technology, (c) to improve your skills in your current job, (d) to prepare you for a job or jobs you might do in the future, (e) to develop your skills generally. It also gets details on training finance by asking who is financing training or exam costs. Finally, it also asks about the duration of three important trainings and whether it has led to a qualification or not. The question in UKHLS waves is stated in this way

In the last 12 months, that is since [interview month] [interview year], have you done any [other] training schemes or courses, even if they are not finished yet? Please include any part-time or evening courses, training provided by an employer, day release schemes, apprenticeships and government training schemes.

Then, it asks about the number of training courses and the provider of three longest training courses with options; (a) provided by employer, (b) government training scheme, (c) college or university degree or diploma course, (d) other type of training course. The training purpose question of UKHLS is not much different from BHPS. Like BHPS, it asks for training duration and whether it has led to a qualification. The training duration question in UKHLS is more precise as it asks number of training hours per day and number of days on training for each of the three longest trainings.

This analysis first focuses on the yes-no answers to the general training question as presented in italic above. Then, it narrows down the training question and presents results for employer provided training and training that had happened at workplace or on-the-job training. For the two-way fixed-effects method, training is not needed to be binary. For this, the training incidence variable is constructed from the question which asks for the number of training courses. Using the permanent human capital formation assumption, training incidence is constructed as a cumulative sum of the number of training courses over time. The two-way fixed-effects estimations also include training intensity variable as well. It is constructed as a cumulative sum of hours spent in the three important training programs.

For the heterogeneity robust methods in Section 4.5, training variable needs to be binary with zero and one values[12]. Keeping this, training incidence measure will take a value of one when the individual is first treated. Given the assumption that training leads to permanent human capital formation (Pischke, 2001), the training variable takes value of one for all periods following the first training year, i.e. $E_i = min\{t : D_{it} = 1\}$. That is, the training variable takes a value of zero for all pre-training periods and value of one for the training year and all years following the training year. This assumption is standard and considered plausible in training literature. But for comparison purpose, this assumption is relaxed later to see the difference in results, as Imai et al. (2018) allow for treatment reversal. Under this alternative, the training variable takes a value of one only for the year where training takes place[13].

The data sets the study starts with in this section are unbalanced panels. It considers only those years for an individual where she is in paid employment; thus

[12]Actually, Callaway and Sant'Anna (2018) and Abraham and Sun (2018) are event study designs, and build on the time of first occurrence of treatment/training.

[13]In Table C.1, all training variables followed by integer 2 are one which are not based on permanent human capital formation assumption. Variables which are based on permanent human capital formation assumption, and are constructed as cumulative sum, are not followed by any integer.

excluding observations if the individual is in self-employment, is in full or part time education or is on maternity leave. Those who are on maternity leave or in schooling are likely to have different wages than usual even if they are getting wages. Similarly, it excludes all years for an individual for which the wage data is missing. After dropping these observations, one is left with unbalance UKHLS panel having 37,417 individuals and 135,405 observations. In the BHPS sample, the total number of observations is 58,199 with 13,299 individuals. The mean values in Table C.1 are presented for these data sets. Also, the two-way fixed-effects estimation results in Section 4.4.3 are based on these data sets. For the matching and weighting estimations, further data restrictions are discussed in Section 4.5.

4.4.2 Descriptive Statistics of Trained versus Untrained

As is clear from Table C.1 in the appendix, the mean number of training courses in the UKHLS sample, including all types of trainings, is 0.80 with a standard deviation of 1.80. In total, 67.22% of the sample undergoes no training annually while 15.41% undergoes one training course. Similarly, 6.99% take two training courses and 4.46% undergo three trainings. Finally, 0.03% of the sample takes thirty training courses. The mean number of employer provided training courses is 0.51. The low mean here is because such information is available for the three important trainings only. In this, 72.55% of the sample has no employer provided training while 12.39% sample undergoes one such training, 6.37% sample takes two employer trainings and 8.69% undergoes three employer trainings annually. The mean number of all training courses among the trained is 1.21 and of employer provided training is 0.78. The mean training hours for all types of trainings are 18.45 with a standard deviation of 78.99, whereas in the trained sample the mean training hours are 27.92 with a standard deviation of 95.80. This shows that most of the trainings in this sample are short training programs lasting for less than a week.

When one assumes that training leads to permanent human capital and follows the cumulative summation approach to measure training, then the mean number of training courses is 2.54 with a standard deviation of 4.75. Out of 37,417 individuals, 46.95% undergo no training in the sample period while 14.45% undergo one training course. Similarly, 9.35% individuals take two training courses and 6.56% undergo three training programs. The mean number of employer provided training courses is 1.64. In this sample, 53.49% individuals undergo no employer provided training while 12.86% undergo one such training program. Similarly, 8.49% take two employer training courses and 9.02% undergo three employer training programs. These statistics show that most of the three reported trainings are employer provided training courses in this sample.

In the BHPS sample, the mean number of training courses, including all types of training, is 0.60 with a standard deviation of 1.37. In total, 70.29% of the sample undergoes no training annually while 16.02% undergoes one training program. Table C.1 also reports training that took place in the workplace for the BHPS panel. It combines the first three training place options, i.e (a) at current or future workplace (b) at former workplace (c) employer's training centre. The mean number of training courses that has taken place at workplace is 0.28. The low mean here is again because such information are available for the three important trainings only. In this, 83.13% of the sample reports no workplace training while 9.34% undergoes one such training course. The mean number of all training courses among the trained is 0.90 and of workplace training for the trained it is 0.42. Most of the reported trainings are on-the-job training courses in BHPS sample. The mean training hours for all types of trainings are 40.92 whereas in the trained sample the mean training hours are 61.40. Thus in BHPS, the mean training hours are greater than in the UKHLS. However, the estimation of training hours is not precise here. This survey asks training duration question as unit of training duration and number of units. The options of units are; (a) hours, (b) days, (c) weeks and (d) months. It is converted into hours by assuming that a day is equal to 5 hours, a week is 5 days and a month is 4 weeks.

The important thing to see from Table C.1 is how other individuals' characteristics differ between the trained and the untrained. Table C.1 reports a number of individual characteristics for the overall sample and the trained sample, which includes all those who got at least one training over the sample period. First, the gross wage of the trained sample is greater than the gross wage of the overall sample in both data sets. The difference implies that if the overall sample's mean wage is £2,000, the trained sample mean wage is approximately £2,200. But this difference may be due to differences in others characteristics of the trained and the untrained rather due to training. Regarding other characteristics, the trained are on average more qualified and work in larger firms in terms of employment size in both data sets. This is according to expectations as a number of studies show that highly qualified get relatively more on-the-job training and larger size firms usually train more (Booth, 1991, OECD, 1991, 1999, Green, 1993, Veum, 1997, Arulampalam and Booth, 2001, Brunello, 2001, Arulampalam et al., 2004).

Second, the mean job tenure and age are both lower in the trained sample than in the overall sample[14]. The finding about age is plausible as mostly young workers are likely to take training. Some expect that training may increase with tenure as longer tenure with a firm increases the chances that the worker will stay, and thus reduces the firms' fear that the worker will leave after training. But given that young workers mostly get training and are more mobile, this finding is not surprising. From the UKHLS data, one can see that the trained are on average in high skilled managerial jobs compared to the untrained; see the mean of job group and manager variables in Table C.1. Although not much different, the trained are more likely to have second job and less satisfied than the untrained. The statistics on job satisfaction are interesting. It hints at the fact that only comparing wage level differences of workers may not be a good way to see how happy they might be with their jobs. The less job satisfaction level of trained may be because they

[14] Job tenure is not reported for UKHLS as it has a lot of missing values and will lead to a significant reduction in the sample size; particularly the approaches used in Section 4.5 do not allow for missing values.

are not rewarded enough for their high skills from education and training. Finally, the trained and untrained are not different consistently along the characteristics of unemployment spells and whether their jobs are permanent or temporary.

4.4.3 Results from the Standard Two-Way Fixed-Effects

In this section, the standard approach used in the training literature during the 1990s and the 2000s is applied to see how the results compare with earlier literature. In the two-way fixed-effects estimations, the time trend is individual specific as specified in Equation (4.1). To check the robustness of results, this study includes the time-varying covariates in steps starting with covariates which are not likely to be affected by training such as age and education. Among time-varying covariates, it includes variables which are included in different earlier studies like job tenure, occupation, industry dummy, job status, unemployment spells and firm size. It also includes variables such as job satisfaction and non-labour income as well. Finally, it includes squares and cubic roots of education, age, job tenure and firm size unless it is not highly insignificant.

The estimation results are given in Tables C.2 and C.3 in the appendix. The dependent variable in Columns 1-5 of each table is log of gross monthly nominal wage. Column 6 presents results for gross real wage. Gross real wage is obtained by deflating nominal wages using UK consumer price index with base year of 2001 for BHPS panel and 2010 for UKHLS data[15]. The first column of each table gives OLS estimation for comparison purpose. As is clear from the tables, the effects of training intensity are close to zero and insignificant. In the BHPS panel, the training intensity coefficient is insignificant under all specifications. In the UKHLS data, it is significant only under the base level specification but with coefficient near zero.

[15]Results are presented for both nominal and real wages. The reasons for this are two-fold. Earlier studies either use nominal or real wage exclusively, but the training coefficient seems different for the two measures of earnings. In this case, training coefficient is high in real wage regression. Second, the coefficients of other regressors in the nominal wage regression is more stable and plausible under different specifications. Since, this study tries to estimate the lower bound/value on training coefficient, it presents results for both nominal and real wages.

After including other time varying variables, it becomes insignificant as well. This finding is in accordance with earlier training literature, which shows that training incidence is important but the duration of the training has no significant effects on wages (Booth, 1991, Arulampalam and Booth, 2001). For example, Veum (1995a) uses the NLSY data for the 1986-1990 period and shows that any form of continuous training measure has no significant effects on wages.

Although training duration seems to have no effects on earnings, participation in training programs has positive and significant effects on workers' earning in all specifications and for both data sets. Column 3 suggests that the effect of a given training incidence is 0.38% in the UKHLS panel and 0.67% in the BHPS panel. Moreover, the two-way fixed-effects estimations significantly reduce the coefficient of training incidence as compared to OLS estimation. This highlights the fact that unobservable individuals specific determinants of wage growth are important and training variable is likely to pick such effects if one fails to control for it. Particularly, controlling for such effects increases the overall R-square from a range of 0.2-0.29 to above 0.9. Comparison of Columns 2 and 3 in each table shows that controlling for time-varying variables also decreases the coefficient on training incidence by a significant margin. Additionally, when one relaxes the assumption of permanent human capital formation effects of training, the coefficient of training incidence decreases in both data sets. The reason for this is that now some of the previously trained individuals are treated as untrained in period t and are included in the control group.

The coefficient of the employer provided training incidence in UKHLS data is high and reaches the OLS coefficient for the overall training. In this case, a given training incidence increases wage by 1.03%. The coefficient of workplace training incidence in BHPS panel is significant but is less than the coefficient on the overall training incidence in Column 3. But this coefficient cannot be compared with the employer provided training in UKHLS data. Employer provided training asks for any training, on or off-the-job, provided by the employer. On the other hand, workplace training is more formally referred to as on-the-job training. Put in this way, this result is expected as the existing literature shows that training return is high for off-the-job training compared to on-the-job training (Lynch, 1991). For example, Blundell et al. (1996) find that on-the-job employer supported training increases wage return by 3.6% for men and has no significant return effects for women. On the other hand, off-the-job training incidence leads to 7% earning increase for men and 5% for women. In this study, the overall training incidence measure includes both on and off-the-job training courses, so its high coefficient as compared to the coefficient of workplace training is not unexpected.

Thus, the coefficient of training incidence varies with a change in the data set, specification and estimation method, and with the definition and construction of training incidence over time. But despites all these variations, training incidence has positive and significant wage effects under the standard fixed-effects estimation. However, the estimated return is low as compared to the return estimates from many studies during the 1990s and the 2000s. But this is not an exception. Parent (1999) estimates training return of 12.16% to a full year of on-the-job training from NLSY panel for the 1979-1991 period. Booth et al. (1993) fixed effects estimation from the British National Survey of Graduates and Diplomats (BNSG) for the period 1986-1987 shows that one week of training in the first year of job increases women earnings by 1%. For Germany, Pischke (2001) uses German Socio Economic Panel (GSOEP) data and estimates Equation (4.1) for the 1986-89 period. In his study, one year work-related training increases earnings in the range of 2.6% to 3.8%. Looking at this and many other studies in this literature (Haelermans and Borghans, 2012), one can conclude that the current study's fixed-effects return estimates to an incidence of training lie below the average return found in the 1990s and the 2000s studies. But this return is not low when one keeps in mind that a given training incidence last for just few hours in these panels.

Regarding the other explanatory variables, except the coefficient of age in the real wage column, coefficients of all the variables have expected signs and remain stable under different specifications. Wage increases with qualifications as expected. Qualifications have cubic effects in the BHPS sample. However, in the UKHLS panel both square and cubic terms of qualifications are insignificant. Firm size has also positive quadratic effects on wage earning in all specifications. Having a permanent job increases workers' wages while holding a second job decreases earnings. Both these results are again quite plausible, as permanent jobs are mostly high paying. Managerial position also has positive and significant earnings effects as expected. Similarly, holding an occupation which requires high skills significantly increases worker's wage. Job satisfaction also has positive and significant wage effects. The likely explanation for this is that efforts may increase with job satisfaction. Finally, unemployment spells during a given year also negatively affects worker's earnings. More unemployment spells during a given time may deteriorate worker skills besides giving negative signals to the employer. In today's world, most employers ask about reason of a job switch, and for any gaps in working history. Moreover, the coefficients of most of the variables in Tables C.2 and C.3 are quite stable across the specifications.

4.4.4 Analysis with Balanced Subpanels

As mentioned earlier, one weak point of experiment data based studies is their sample size. The sample size is important in the case of training because, unlike formal schooling, there are a lot of types and arrangements under which training can take place. This can create a significant variation in the wage return of training. To see how the sample size changes affect the significance of training coefficient, this section performs the estimations with subsamples. For this purpose, it takes balance subpanels from each data set. The results are shown in Table 4.1. It starts with only those in the first row for whom data is available for two years only. Then, it considers those with three years data and so on up until those with seven years of data. To avoid differences due to model specifications, the estimation is carried with exactly the same specification under all subsamples.

As is clear from Table 4.1, there is a very clear negative relation between the sample size and standard errors in both data sets. In BHPS, the standard error is 0.008 with a sample size of 3,100, and it drops down to 0.002 when the sample size has increased to 9,391. The negative relation is even stronger in the UKHLS data set. Moreover, as the sample size increases the coefficient of training becomes significant in the BHPS balance panel. In the UKHLS sample, it is insignificant for most of the cases. But in the subsequent analysis in Section 4.5, the study shows that the UKHLS data does not satisfy the underlying assumption of common parallel trends. The BHPS panel satisfies the parallel trends assumption, so these results seem more reliable. Additionally, the size of the training coefficient is on average the same in subsamples as it was in the estimates from the overall sample. The basic message this exercise gives is that the sample size can play important role in the significance level. The studies conducted in the 1990s and the 2000s were based on mutli-periods panel data sets and their rich sample size has allowed for precise estimation. On the other hand, the studies in the 2010s are mostly based on two periods and have a small sample of observations, as discussed in Section 4.2. Thus, the insignificant results in such studies might not be a strong conclusion. In the next section, the study considers the question of heterogeneity in the effects of training.

Table 4.1: Two-Way Fixed-Effects Results from Balanced Subpanels

Sample	*Sample Size*	*BHPS*	*Sample Size*	*UKHLS*
Two Periods	3,100	0.0040	6,084	0.0077
		(0.0076)		(0.0063)
Three Periods	3,563	0.0058	10,263	0.0060
		(0.0046)		(0.0033)
Four Periods	4,267	0.0153	12,030	0.0051
		(0.0038)*		(0.0024)***
Five Periods	4,633	0.0124	14,111	0.0024
		(0.0036)**		(0.0015)
Seven Periods	9,391	0.0055	45,502	0.0003
		(0.0019)**		(0.0005)

Note: This table shows results for training coefficient only.*,**,*** show significance at 1%, 5% and 10% levels, respectively. It includes all explanatory variables which were included in the full sample estimates, except training hours. Standard errors are given in parenthesis.

4.5 Results from the Weighted DID and Refined Fixed-Effects Methods

This section presents results for training return from the application of Callaway and Sant'Anna (2018), Imai et al. (2018), and Abraham and Sun (2018). As discussed above, Abraham and Sun (2018) and Callaway and Sant'Anna (2018) offer techniques that address the problem of possible negative weights when treatment effects are heterogeneous. Imai et al.'s (2018) method also controls for the history of past wages in order to purge the effects of past outcomes on the current treatment.

Like all other matching and weighting methods, these approaches put certain restrictions on the data. The basic purpose behind such restrictions is to control for the differences in observed characteristics between the treated and untreated units. First of all, DID requires that no one should be treated in the first or pre-treatment period. Furthermore, for a formal test of pre-treatment parallel trends, one should have at least two periods pre-treatment data on each of the unit. Thus, individuals who are trained in the first two periods are dropped in the application of Callaway and Sant'Anna's (2018) method. Secondly, it considers only those individuals for whom at least three years of data, two pre-training and at least one training year, is available. Finally, this method does not allow for missing values on any of the pre-treatment covariates. After dropping these observations, one is left with 21,395 observations with 3,228 individuals in the BHPS panel. The UKHLS data has 6,440 individuals with 36,664 observations.

For Imai et al.'s (2018) technique, one does not need to drop observations manually. Since it matches individuals on the basis of past treatment history, past outcomes and time-varying covariates, it drops observations accordingly. For example, if one continues with the assumption of permanent human capital formation and keeps lag value of two, then it will drop all the individuals who are treated in the first two periods. Besides this, it will drop trained individuals who cannot be matched on either past outcomes or on time-varying covariates. Additionally, it allows one to choose the number of maximum matched control units for each treated unit. Thus, it will carry out estimation using only the sets of those control matched units.

For the Abraham and Sun's (2018) application, this study considers only balance panel from each data set. It further restricts the analysis to consider only those who got at least one training course during the sample period. The purpose of this is to ensure that the observations are as similar as possible. Thus, in the application of this method, the control units are the pre-training observations of the trained individuals. This must make the control and trained groups strongly similar. After these restrictions, the BHPS panel has 10,208 observations whereas the UKLHS has 17,087 observations for the Abraham and Sun (2018) application. For each of the reduced panel, the training incidence coefficient from the standard fixed-effects method is also reported for comparison.

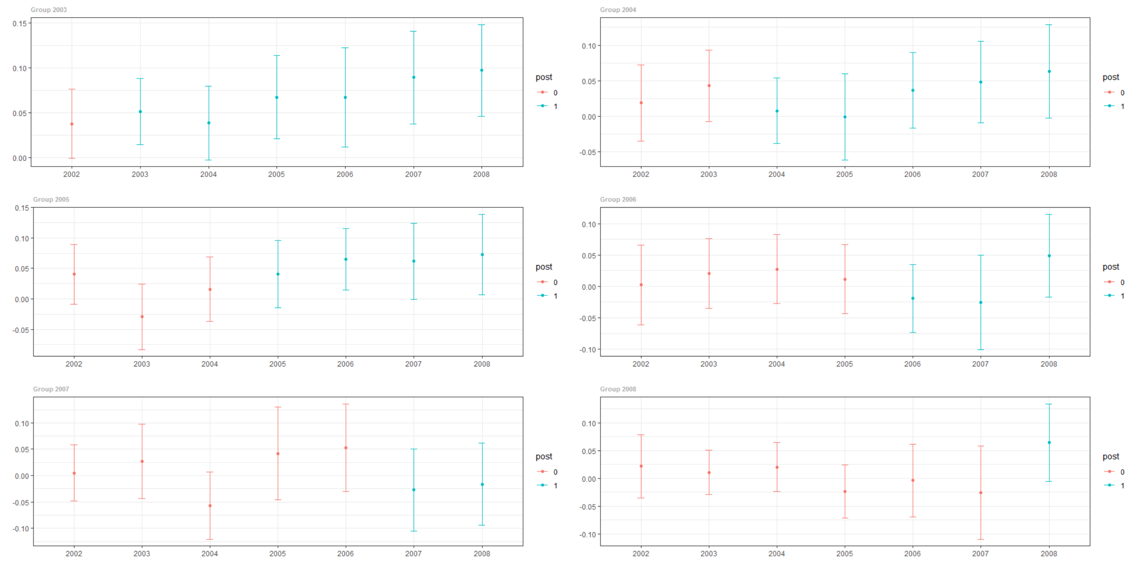

Figure 4.1: Un-Conditional DID Estimations Based on BHPS

4.5.1 Results from Callaway and Sant'Anna (2018) Application

As highlighted in Section 4.3, this method calculates group-time average treatment for the treated $ATT(e,t)$. Then, one can aggregate it into a single parameter. For discussion sake, some of the group-time $ATT(e,t)$ are given in Figures 4.1-4.3 for the BHPS panel. Individual group-time $ATT(e,t)$ for the UKHLS panel are given in the appendix. The vertical line in the figures shows 95% point-wise confidence band. The red lines give pre-training $ATT(e,t)$ and is a source to check the validity of the parallel trends assumption. If the pre-training parallel trends assumption is true, then the pre-training $ATT(e,t)$ shown in the red lines should not be significantly different from zero.

All the aggregated results are presented in Table 4.2. Column 1 of the table presents results for nominal wage under unconditional common trends for each data set. Aggregate $ATT's$ for real wage are given in Column 2 for each data. As is clear from the table, the aggregate earning effects are low for the real wage estimations. Keeping this, rest of the columns report estimations with the real wage. The aggregate real wage effects of training under conditional common trends are given in Column 3 for each data. These aggregates are estimated from group-time $ATT(e,t)$

in Figures 4.2 and C.2, respectively. For the estimations under conditional parallel trends, this study controls for pre-training values of all the variables used in the two-way fixed-effects estimations in Tables C.2 and C.3. Columns 4 carry estimations on the reduced sample of those who experienced training incidence in only one year over the sample period, to see whether training has true dynamic effects. Finally, the clustered bootstrapped standard errors at the individual level are given in the parenthesis.

First, note from the figures and p-values of Wald test in Table 4.2, that both the unconditional and conditional parallel trends assumptions seem to fail in the UKHLS panel. This is true under both point-wise and uniform confidence bands. This is particularly clear from looking at the groups first treated in 2015 and 2016 in Figure C.2. For these groups, the $ATT(e, t)$ fluctuate and are far away from zero. On the other hand, in the BHPS panel, the assumption of pre-training common trends cannot be rejected. When conditioned on the pre-training covariates, the p-value of Wald test for the null that pre-training $ATT(e, t) = 0$ is 0.44. Given that pre-training parallel trends assumption cannot be rejected for the BHPS data, the study focuses on these results.

Regarding the effects of training, from Figure 4.1, the $ATT(e, t)$ of training estimated under unconditional parallel trends steadily increase over time. Seven of the post-training $ATT(e, t)$ are statistically significant at 5% significance level. In terms of magnitude, the effects of training under this application are higher than the results in fixed-effects estimations. For example, for the group treated in 2003, its 2008 average earnings are 9.7% higher as compared to it would have been under no training. For the group treated in 2004, the nominal $ATT(e, t)$ are 6.3% higher in the year 2008. Similarly, the group treated in 2005 has a maximum pre-training $ATT(e, t)$ of 0.04 in 2002. However, in the post-training period, its $ATT(e, t)$ reach 0.073 in the year 2008. Moreover, many of the post-training $ATT(e, t)$ are significant at 10% significance level.

Table 4.2: Aggregate Training Effect Estimates with Callaway and Sant'Anna (2018)

Aggregation	BHPS Panel				UKHLS Panel			
Method	1	2	3	4	1	2	3	4
Simple ATT	0.049	0.039	0.031	0.029	0.028	0.017	0.016	0.023
	(0.015)*	(0.013)*	(0.014)**	(0.018)	(0.010)*	(0.009)***	(0.009)***	(0.011)**
Selective ATT	0.043	0.034	0.027	0.024	0.026	0.017	0.015	0.023
	(0.015)*	(0.013)*	(0.014)***	(0.018)	(0.011)**	(0.009)***	(0.009)***	(0.011)**
Dynamic ATT	0.057	0.045	0.036	0.038	0.035	0.022	0.019	0.025
	(0.015)*	(0.014)*	(0.015)**	(0.022)***	(0.012)*	(0.011)***	(0.010)***	(0.014)***
Calendar ATT	0.047	0.039	0.032	0.032	0.023	0.013	0.012	0.024
	(0.013)*	(0.012)*	(0.013)**	(0.020)	(0.009)**	(0.009)	(0.009)	(0.013)***
Test of DID	p=0.28	p=0.40	p=0.44	p=0.61	p=0.04	p=0.04	p=0.01	p=0.11

Note: Significance at 1%, 5% and 10% levels is indicated by *, **, ***, respectively. The last row reports p-value of Wald-test for pre-training common trends assumption. Standard errors are estimated with 1000 bootstrap iterations. The two-way fixed-effects coefficient estimated for this sample of BHPS is 0.009 for nominal and 0.006 for real wage. Columns 1 and 2 are based on unconditional pre-training common trends assumption. Columns 3 and 4 are based on conditional common trends. Columns 4 estimate *ATT* for those who experienced training in a single year only to try estimate dynamic effects.

Looking at the aggregate effects, one can see from Columns 1 and 2 of Table 4.2 that the aggregate effects of training, estimated under un-conditional common trends, are positive and significant no matter which method of aggregation one considers. From these columns, one can also see that the training effects on real earnings are bit low in magnitude as compared to nominal wage effects. The aggregate real earning effects of training are 3.9% under the simple aggregation and 4.5% under the dynamic aggregation. Moreover, under all the specifications, the aggregate earnings effects are higher under the dynamic aggregation. This should be the case as is clear from the figures. The groups treated in 2003 and 2004 are having high $ATT(e,t)$ as compared to other groups. Under dynamic aggregation, these groups get more weights thus leading to a high overall effect of training.

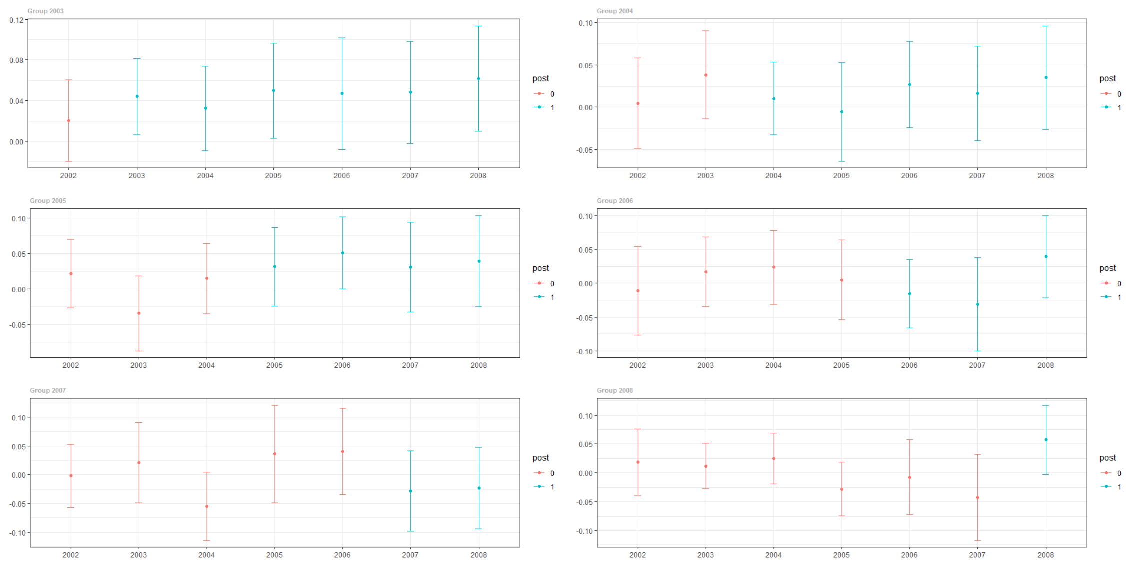

Figure 4.2: Conditional DID Estimations based on BHPS

From Figure 4.2, the $ATT(e,t)$ of training on the real wage with the conditional parallel trends assumption are smaller as expected. This means that at least some of the wage differences between the trained and untrained are due to the differences in their observed characteristics. This should be the case as we saw for example that the trained are also more qualified as well. Now, four of the $ATT(e,t)$ are significant at 5%. In terms of magnitude, the aggregate real earning effects of training are now 3.1% under the simple aggregation and 3.6% under the dynamic aggregation. Thus, even after controlling for pre-training characteristics, the wage effects of training are higher from the application of Callaway and Sant'Anna's (2018) weighted DID as compared to two-way fixed-effects. In the UKHLS panel, only one $ATT(e, t)$ is significant after controlling on pre-training covariates. Also, the aggregate earning effects of training are lower in this panel. For example, the aggregate real earning effects with the dynamic aggregation are 1.9% in this panel. However, the pre-training parallel trends assumption does not hold for this panel even after controlling for the observed characteristics.

Two points are important about the above results. Firstly, the recent findings that in presence of heterogeneous treatment effects, the two-way fixed-effects can assign negative weights seem important. One can see in the figures that we have both positive and negative $ATT(e, t)$ in the post-training period. Moreover, the $ATT(e, t)$ are different for the different groups and change over time; clearly indicating heterogeneity of treatment effects. From Figure 4.2, five of the post-training $ATT(e, t)$ are negative. If these get negative weights in the fixed-effects method, they will be counted positive in the aggregation as is clear from Equation (4.2). On the other hand, if the positive $ATT(e, t)$ get negative weights these will be subtracted in the aggregation process. This means that the wage effects of training could either be over or under estimated. Secondly, there is an upward trend in the post-training $ATT(e, t)$. This trend is more obvious the longer is the post-training period. This result hints at the possibility that trained workers reap the benefits of training in the long run.

However, interpretation of the upward trends in $ATT(e, t)$ from the first three columns of each panel as dynamic effects of training is not without problem. The high $ATT(e, t)$ for the later periods might be due to the fact that the trained are treated repeatedly following their first training, and the training dummy picks the effects of these subsequent periods trainings. To check for this possibility, the panels are reduced by dropping all those who are trained in more than one year. The results for these subsamples are presented in Columns 4 of Table 4.2. The corresponding

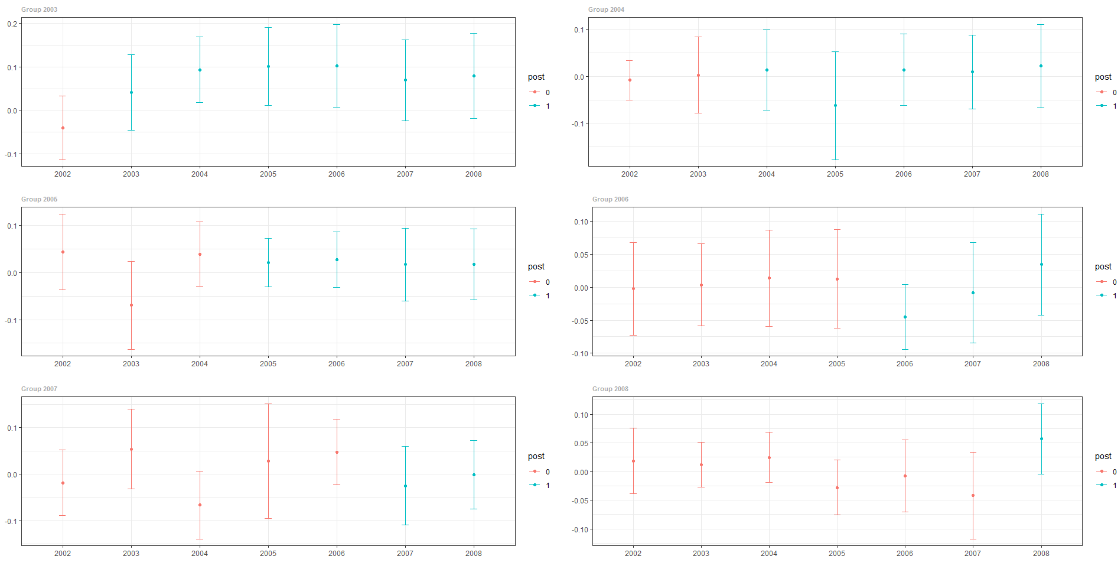

Figure 4.3: Conditional DID Estimations for a single year Training from BHPS

$ATT(e,t)$ for the BHPS panel are given in Figure 4.3. These estimations are carried with the same specification as for the conditional DID in Column 3. The pre-training common trends assumption holds for both panels. This result is not very different from the results in Column 3. Only the significance level drops down in the BHPS panel. This is because of the big drop in the sample size. The dynamic aggregate real wage effects are even higher in these estimations as compared to the results in Column 3. This hints at possible dynamic effects of training over many future periods[16]. To more better match the treated and control groups, the study now proceeds to the application of Imai et al.'s (2018) method in the next subsection.

4.5.2 Results from Imai et al. (2018) Application

To allow for enough lags and leads, Imai et al.'s (2018) approach needs panel with a relatively longer time span. Thus, it is only implemented on the BHPS sample. The results are presented in Table 4.3. All the estimations use Mahalanobis refinement method except Column 2. Refinement in Columns 1 to 4 is based on past wages only, using two lags. Column 3 presents results when the assumption of permanent human capital formation of training is relaxed. This column carries estimation for

[16]This question will be addressed again in the application of the other approaches.

two post-training periods only due to the short time span of the data. Column 4 reports estimation results with a matched set size of 10 to see if there is any significant difference in the results when the match size changes. Column 5 presents results when job tenure, qualifications, firm size and age are included as time-varying covariates besides past wages. Column 6 includes additional covariates like dummies for the second job, permanent job, marital status, and the non-labour income. Column 7 of Table 4.3 presents results for the workplace training and Column 8 reports estimation for real wage. Column 9 carries estimation on a reduced sample of those who experienced training incidence in only one year, to see whether training has true dynamic effects.

As is clear from Table 4.3, training has positive and highly significant effects on the gross monthly wages of the trained[17]. The positive and significant wage effects of training hold for both refinement methods and under different specifications. First, the $ATT's$ are high when propensity score matching is used for the refinement. This is evident from the comparison of Columns 1 and 2 of the table. Given this, Mahalanobis matching method is used for rest of the estimations to get the lower possible $ATT's$ within this setup, as the study tries to get lower bound of the training return. Secondly, comparison of Columns 1 and 4 of the table shows that increasing the number of control units in the matched set does not change the result by a significant margin. Given this, a match set size of 3 is used in rest of the columns to keep machine memory requirements and iterations timing not very involved. Third, the introduction of the first four covariates decreases the effects more than the introduction of all eight covariates. Fourth, as is clear from Columns 6 and 8 comparison, there is no significant difference between estimations for real and nominal wages.

Like in Callaway and Sant'Anna (2018), the effects of training are higher in the long run. As is clear from Table 4.3, the $ATT's$ increase over time in all the specifications. The $ATT's$ in periods $t+3$ and $t+4$ are far higher than the immediate

[17]For graphical view of the results in Table 4.3, see Figures C.3-C.10 in the appendix.

Table 4.3: Training Effect Estimates with Imai et al. (2018)

Timing	1	2	3	4	5	6	7	8	9
$t+0$	0.059	0.063	0.038	0.054	0.032	0.047	0.042	0.047	0.060
	(0.015)*	(0.019)*	(0.010)*	(0.015)*	(0.016)	(0.015)*	(0.021)	(0.015)*	(0.041)
$t+1$	0.038	0.070	0.045	0.041	0.037	0.034	0.037	0.035	0.078
	(0.020)	(0.020)*	(0.011)*	(0.017)*	(0.018)*	(0.019)	(0.025)	(0.019)	(0.041)
$t+2$	0.084	0.097	0.067	0.074	0.059	0.062	0.045	0.061	0.127
	(0.022)*	(0.023)*	(0.013)*	(0.020)*	(0.020)*	(0.022)*	(0.028)	(0.022)*	(0.044)*
$t+3$	0.081	0.104	-	0.073	0.061	0.065	0.075	0.063	0.109
	(0.024)*	(0.025)*	(-)	(0.022)*	(0.025)*	(0.025)*	(0.030)*	(0.028)*	(0.046)*
$t+4$	0.088	0.142	-	0.089	0.071	0.083	0.105	0.079	0.094
	(0.025)*	(0.025)*	(-)	(0.022)*	(0.024)*	(0.026)*	(0.032)*	(0.031)*	(0.047)
lags (L)	2	2	2	2	2	2	2	2	2
Match Size	3	3	3	10	3	3	3	3	3
No. of Covar.	$Wage$	$Wage$	$Wage$	$Wage$	$Wage+4$	$Wage+8$	$Wage+8$	$Wage+8$	$Wage+11$
Refinement.	Maha	PS	Maha	Maha	Maha	Maha	Maha	Maha	Maha

Note: Covar. denotes number of covariates. Maha means refinement through Mahalanobis and PS stands for propensity scores. Significance at 5% level is denoted through *. Standard errors are estimated with 1000 bootstrap iterations.

effects of training. However, this interpretation is once again not definitive until unless one controls for the fact that some individuals are trained in more than one period. To check for this possibility, all those are dropped once again who are trained in more than one year. The result for this subsample is presented in Column 9 of the table. With this reduced sample, covariates like job satisfaction, managerial duties and unemployment spells are also used for refinement of the matched set[18]. As is clear from Column 9, the $ATT's$ of training still increase over time. The $ATT's$ of training are higher in years 3 and 4 as compared to the immediate training year. The standard errors are now higher, but this is due to reduced sample after dropping many observations. Combined with similar findings from Callaway and Sant'Anna (2018), these results clearly show that workers are getting relatively more wage benefits of training in the long run. This result adds strength to the belief that training has dynamic effects, as here the study controls for eleven time-varying covariates' current and past values besides past wages and training history. This finding also justifies the assumption that training has human capital formation effects over many years.

Another aspect of Table 4.3 is that it gives relatively larger effects of training on wages of the trained. The minimum ATT it gives is 3.2% in the training period, i.e. $t + 0$. On the other hand, even after controlling for many covariates' history, the $ATT's$ in the fourth year after the training $(t + 4)$ lie in the range of 7.1%-14.2%. However, when comparing to the fixed-effects and the results from Callaway and Sant'Anna (2018), one must keep in mind that the underlying sample is not exactly the same for the Imai et al. (2018) application. Finally, comparison of Columns 6 and 7 of Table 4.3 suggests that the immediate effects of the workplace training is lower than the effects of the overall training measure. But in the later periods, the effects of workplace training gotten greater than the effects of the overall training. This result holds even if one adds the number of training hours among the covariates. The lower ATT of workplace training for $t+0$ thus matches the two-way fixed-effects

[18]Note that the results in Column 9 hold with covariates set as in earlier columns.

results. But, it gives the additional insight that workplace training wage effects are likely to overweight off-the-job training effects in the long run. However, to reach a conclusion about these findings, one needs more careful analysis.

Table 4.4: Training Effect Estimates with Abraham and Sun (2018) from BHPS Subpanel

	FE	IWD	Estimates for ATT for the six cohorts						IWD
r'	$Est.$	$Est.$	Coh_1	Coh_2	Coh_3	Coh_4	Coh_5	Coh_6	$Est.$
-5	-0.025	-0.011	-	-	-	-	-	-0.03	-0.023
	(0.045)	(0.016)						(0.04)	(0.021)
-4	-0.030	-0.002	-	-	-	-	-0.02	0.00	-0.001
	(0.081)	(0.005)					(0.04)	(0.05)	(0.011)
-3	-0.048	-0.003	-	-	-	0.02	-0.01	-0.08	-0.006
	(0.114)	(0.009)				(0.04)	(0.05)	(0.05)	(0.016)
-2	-0.059	-0.003	-	-	0.00	-0.02	0.00	-0.02	-0.013
	(0.149)	(0.015)			(0.04)	(0.04)	(0.06)	(0.07)	(0.024)
-1	-0.043	0.013	-	0.02	0.04	-0.02	0.04	0.04	-0.005
	(0.183)	(0.022)		(0.03)	(0.04)	(0.05)	(0.06)	(0.07)	(0.030)
0	-0.033	0.031	0.01	0.07	0.03	0.04	0.02	0.04	0.003
	(0.217)	(0.032)	(0.03)	(0.04)**	(0.05)	(0.05)	(0.06)	(0.07)	(0.037)
1	-0.040	0.036	0.03	0.03	0.04	0.06	0.04	-	0.001
	(0.252)	(0.036)	(0.04)	(0.04)	(0.05)	(0.05)	(0.07)		(0.040)
2	-0.025	0.053	0.01	0.08	0.08	0.08	-	-	0.023
	(0.285)	(0.039)	(0.04)	(0.05)	(0.05)	(0.06)			(0.043)
3	-0.031	0.073	0.05	0.08	0.12	-	-	-	0.043
	(0.319)	(0.042)**	(0.05)	(0.05)	(0.06)*				(0.046)
4	-0.027	0.082	0.05	0.13	-	-	-	-	0.045
	(0.354)	(0.046)**	(0.05)	(0.06)*					(0.054)
5	-0.019	0.106	0.11	-	-	-	-	-	0.079
	(0.388)	(0.056)**	(0.06)**						(0.070)
6	-0.018	-	-	-	-	-	-	-	-
	(0.422)								
Obs.	10,208		8,932	8,932	8,932	8,932	8,932	8,932	
Adj.R^2	0.8542		0.8542	0.8542	0.8542	0.8542	0.8542	0.8542	
W. R^2	0.2073		0.1824	0.1824	0.1824	0.1824	0.1824	0.1824	

Note: Only those individuals are part of this subpanel who have data for all waves and are trained at least once in the sample period. Significance at 5% and 10% levels is indicated by * and **, respectively. Adjusted and within R-squares are given in the last two rows, respectively. The standard two-way fixed-effects training incidence coefficient for this sample of 10,098 is 0.003 (0.003) and is insignificant with individual specific time trends, but with common time trend it is 0.004 (0.002) and is significant.

4.5.3 Results from Abraham and Sun (2018) Application

Imai et al.'s (2018) method is more general and addresses many of the possible doubts. One thing that is not accommodated in matching or weighting DID methods is unobserved unit-specific effects. On the other hand, such a capability of fixed-effects regression is one major reason behind its prevalence. Fortunately, Abraham and Sun's (2018) approach allows for unit and time fixed-effects and at the same time addresses the earlier highlighted issue of the standard fixed-effects; that is, it is not robust to heterogeneous treatment effects. To overcome the doubt that the above results might be due to the unobserved differences among the trained and untrained, this section also replicates Abraham and Sun (2018). To further improve the reliability of the results, in these estimations only those individuals are considered who are trained at least once during the sample period. This approach is applied to both the BHPS and UKHLS panels. This data has a big sample size and one is left with enough observations even after dropping all the never-trained individuals. Furthermore, it keeps balanced panels of all those who have data for all waves of the respective panel.

The results for this are shown in Table 4.4 for the BHPS subpanel. In this table, the second column gives dynamic two-way fixed-effects as in Equation (4.14). The IWD estimates in Column 3 are the weighted averages of $CATT(e, r')$ and are estimated from Equation (4.16). Weights are the sample shares for all those cohorts which are treated at least r' periods. In Table 4.4, Cohort 1 in Column 4 consists of those who got trained in the year 2002, whereas Cohort 6 in Column 9 are those trained in the year 2007. From the table, one can see that Cohort 6 has no $CATT(e, r')$ in post-training time because it is trained in 2007, which is the last year for the analysis. Thus, for this wave one can get only the contemporaneous $CATT(6, 0)$. On the other hand, Cohort 1 got first training in the year 2002, so one can estimate five years post-training $CATT(e, r')$ for this cohort. Table C.4 in the appendix has the same explanation with the only difference that it has five cohorts with Cohort 1 treated in wave 3 of UKHLS and Cohort 5 treated in wave 7. The last column in Table 4.4 re-estimates IWD from the reduced panel of those who got training in a single year only to check the dynamic effects of training again. For this, the cohort specific results, $CATT(e, r')$, are not reported.

First of all, in the UKHLS panel, the null of pre-training parallel trends is rejected. Both, testing IWD coefficients or $CATT(e, r')$ against zero for all $r' < 0$ shows that the null can be rejected at a p-value of ≤ 0.031. This result is obvious from the $CATT(e, r')$ of Cohort 4 for $r' < 0$ in Table C.4. This finding confirms the earlier result that the conditional parallel trends assumption does not hold in the case of UKHLS panel. In the case of BHPS panel, one fails to reject the null of parallel trends. Thus, the study once again focuses on the results from BHPS in Table 4.4. First, note from Table 4.4 that the FE substantially underestimates the return to training. For all $r' \geq 0$, the FE coefficients are negative. On the other hand, there is no single negative $CATT(e, r')$ for all $r' \geq 0$ in Columns 4-9. This confirms the recent concerns that the standard fixed-effects can assign negative weights during aggregation if the treatment effects are heterogeneous. Since the cohort specific $ATT's$ are positive in this case, it implies that some of the $CATT(e, r')$ get negative weights in the FE estimation, leading to an under estimation of the true training return. Second, from the $CATT(e, r')$ in Columns 4-9 and the IWD estimates for $r' \geq 0$, the training return steadily increases in the post-training periods. Fore example, the contemporaneous effect of training, i.e. $r' = 0$, is 3.1%. Then, it increases steadily overtime and reaches 10.6% at the fifth lead. Most of the cohort-specific effects also increase in a similar fashion.

To check for true dynamic effects, once again all those with training incidence in more than one year are dropped from the BHPS panel. The results are shown in the last column of Table 4.4. This once again shows that training increases earning more in the long run. The results are not significant in this column as this estimation put strong restrictions on the data, which has resulted in a sample size of 3,480 for this estimation. Note that with such a low sample size, the fixed-effects coefficient was also insignificant as shown in Table 4.1. Furthermore, the overall significance in this approach is less than the earlier approaches. The reason may be that it includes many leads and lags besides fixed-effects. Besides this, it is well known in the literature that the standard fixed-effects regression estimates the coefficients more precisely than the matching and weighting methods (Callaway and Sant'Anna, 2018). Keeping this in view, the insignificant estimates in some of the subpanel should not be a concern. Finally, like Imai et al. (2018), the training return here is far higher than the one in the standard fixed-effects estimations. This hints at the possibility that the standard fixed-effects regression under estimates the true return to training in presence of heterogeneity.

4.5.4 Sum up

This section summarises some of the main results in Table 4.5. In this table, the underlying sample is the same for fixed-effects and Abraham and Sun's (2018) method application. But it is different for the Imai et al. (2018) application. When no leads or lags of training are included, the fixed-effects coefficient of an incidence of training is 0.004 from this sample. On the other hand, the current and leads fixed-effects coefficients are negative and insignificant when one adds lags and leads dummies relative to the first training event. As noted by Abraham and Sun (2018), this shows that in the presence of heterogeneity, introducing lags can aggravate rather than purging the effects of past treatment.

The results from the weighted DID method of Imai et al.'s (2018) and refined fixed-effect method of Abraham and Sun's (2018) are close to each other and give a similar message[19]. Both these show that training has wage effects which are higher in magnitude than the one shown in fixed-effects estimations. Second, the wage effects of training increase overtime in the post-training period. According to both these approaches, most of the training return is realized after the second year of

[19]Note that results from Callaway and Sant'Anna (2018) show similar pattern but slightly small in magnitude. Moreover, IWD results from Abraham and Sun (2018) are not mostly significant. But this is because of stringent restrictions on the sample and including many leads and lags besides unit and time effects.

training. Thus, although the fixed-effects estimates are low from the British Household Panel Survey, the wage return estimates of training from the heterogeneity robust techniques are equal to the one found in studies during the 1990s and the 2000s.

Table 4.5: Results Summary

Time after first training	FE $Est.$	Multi-periods Training IWD	Multi-periods Training Imai et al. (2018)	Single-period Training IWD	Single-period Training Imai et al. (2018)
0	-0.033	0.031	0.047	0.003	0.060
	(0.217)	(0.032)	(0.015)*	(0.037)	(0.041)
1	-0.040	0.036	0.034	0.001	0.078
	(0.252)	(0.036)	(0.019)***	(0.040)	(0.041)***
2	-0.025	0.053	0.062	0.023	0.127
	(0.285)	(0.039)	(0.022)*	(0.043)	(0.044)*
3	-0.031	0.073	0.065	0.043	0.109
	(0.319)	(0.042)***	(0.025)*	(0.046)	(0.046)*
4	-0.027	0.082	0.083	0.045	0.094
	(0.354)	(0.046)***	(0.026)*	(0.054)	(0.047)***
5	-0.019	0.106	-	0.079	-
	(0.388)	(0.056)***		(0.070)	

Note: Significance at 1%, 5% and 10% levels is indicated by *, ** and ***, respectively. The standard two-way fixed-effects training incidence coefficient for the FE sample is 0.0042 (0.0017) and is significant.

4.6 Conclusion

This study is motivated by two findings. One of the finding is specifically from the literature of job related training and another is related to econometric theory. First, most studies during the 1990s and the 2000s that used fixed-effects linear regression models, estimate positive and very high wage return of training (Blundell et al., 1999, Parent, 2003, Arulampalam and Booth, 2001, Frazis and Loewenstein, 2005). On the other hand, later studies give insignificant wage return of training from random experiments data. Second, recent cohort of econometric theory papers put in doubt the fixed-effects estimations itself when applied to data with more than two periods. These studies show that the standard fixed-effects methods are not robust to treatment heterogeneity in multi-periods panel data sets. In the presence of treatment heterogeneity, the standard fixed-effects method is likely to assign negative weights when aggregating different treatment effects into a single parameter.

Given these findings, this study runs the newly proposed heterogeneity robust techniques on multi-periods British Household Panel Survey to re-estimate the true wage return of job training. Given that each technique has its own merits and demerits, this study applies three different techniques to address many of the possible doubts about the true casual effects of training. All three approaches give similar results that are summarized as; (1) training has positive and significant effects on workers' wages, (2) in terms of magnitude, the wage return to training in a given year is at least 3%, (3) the wage return from the heterogeneity robust approaches is greater than the wage return from standard two-way fixed-effects estimations, (4) the return of training persists for many years and picks in later years relative to the contemporaneous effects of training, and (5) the high return remains irrespective of whether one considers on-the-job or off-the-job training.

The conclusion one can draw from this exercise is that training has significant long term wage effects. However, the random experiments based studies during the 2010s mostly use single immediate post-training period data for their analysis. In such an exercise, the long term wage effects of training cannot be estimated properly. Secondly, there is a lot of variation in the types and nature of training, and arrangements under which training takes place. This is likely to create significant variations in the wage return of training. So, to estimate a precise return of training, one needs large sample size. However, in the experiment based studies, the sample size is relatively small. Thus, it is highly desirable to have follow-up surveys of these random experiments in order to correctly and precisely estimate wage return of training from these voucher programmes. Furthermore, this exercise shows that the standard fixed-effects methods can be problematic in the presence of treatment effects heterogeneity.

Chapter 5

Conclusion

This thesis has studied three topics in the field of human capital accumulation and wage setting process. Each chapter constitutes a self-contained paper which explores a different topic but each one is related somehow to on-the-job human capital accumulation.

Chapter two reported evidence from existing studies which shows that firms are charging rent on capital, training and R&D investment. However, the result about whether firms charge rent on working hours is ambiguous in these studies. The chapter then used firm-level panel data of Belgium firms from Konings and Vanormelingen (2015) and showed that Belgium firms do not charge any rent on working hours even though they charge rent on capital and workers' training. The chapter goes on to show that the existing non-competitive models of human capital accumulation cannot explain this empirical finding if wage is set through bargaining process and/or workers' quit decision is exogenous as most of these models assumed. To explain the evidence, this chapter developed a model where the firm invests in both firm-specific and general trainings and the worker takes optimal quit decision but has no bargaining power. It has shown that in such a setting the firm optimal wage strategy is to charge rent on any factor which is under firm's control like training and capital. On the other hand, the firm does not charge rent on factors that are decided by the worker such as work hours, job efforts etc.

This model is also a contribution to the literature which tries to address the question of why firms are financing workers general training even if worker is freely mobile. The two dominant theoretical explanations for firm's such investment incentives are based on firm-specific skills and asymmetric information, respectively. This chapter relaxed some of the assumptions of the models that emphasize the role of firm-specific skills in firm's investment incentives in general skills. It also presented evidence which is indicative of the possibility that the joint existence of firm-specific and general transferable skills can better explain the firm's investment incentives in workers' general skills as compared to explanations that are based on asymmetric information among firms.

The third chapter addressed the empirical finding that more educated workers get more training during job and its implications for wage changes. It developed a heterogeneous agents macroeconomics model where workers invest time and goods in human capital. The results showed that an individual training time increases in pre-job schooling given that pre-job schooling improves the efficiency of training time in future human capital generation or affects individual's preferences. But the preference effects of schooling alone do not create plausible consumption, human capital and wage distributions. Under the assumption that pre-job schooling has a direct role in human capital accumulation, the model generated wage distribution is increasing in schooling and can explain substantial portion of the US workers' median earning distribution by education categories. This setting has predicted that highly qualified individuals spend the time they get from working less hours in learning instead of leisure. This prediction is in contradiction of the standard labour-leisure choice model but seems more realistic and needs empirical investigation.

The fourth chapter has investigated wage return from training by applying the recently developed heterogeneity robust techniques on two sets of British Household Panel Surveys data. Studies during the 1990s and the 2000s that used panel data from household surveys give wage return that is very high. But randomised experiments based studies in the 2010s estimate low and insignificant return from training. The chapter has shown that though the standard fixed-effects methods used in earlier studies are not robust to heterogeneity in treatment effects, but the return from training is positive and significant. The immediate post-training wage effects of training are low and mostly insignificant. But the return from training increases in the long run. Thus, training seems to have dynamic effects on the wage return for many years. The chapter concluded that the insignificant return estimates from studies in the 2010s are due to the fact that these studies use immediate post-training period data only to estimate training return, and have small sample sizes for the analysis.

Appendix A

Appendix to Chapter 2

A.1 Proof of Lemma 1

Proof. To prove Lemma 1, differentiate (2.22) with respect to n_i to get $\frac{\partial W_2^*(T_g,T_s,n_i,n_e)}{\partial n_i} = \frac{\frac{\partial f(T_g,T_s,n_i)}{\partial n_i}}{2+\frac{(1-D(.))\acute{d(.)}}{[d(.)]^2}} + \frac{[1+\frac{(1-D(.))\acute{d(.)}}{[d(.)]^2}]\frac{\partial C_i(n_i)}{\partial n_i}}{2+\frac{(1-D(.))\acute{d(.)}}{[d(.)]^2}}$. This is a continuous function given the assumptions of the model including continuous differentiability assumption on $D(.)$. Furthermore, from Assumptions 1 and 2, $\lim_{n_i \to 0} \frac{\partial W_2^*(T_g,T_s,n_i,n_e)}{\partial n_i} = \infty$ and $\lim_{n_i \to \bar{n}_i} \frac{\partial W_2^*(T_g,T_s,n_i,n_e)}{\partial n_i} = \lim_{n_i \to \bar{n}_i} \frac{[1+\frac{(1-D(.))\acute{d(.)}}{[d(.)]^2}]\frac{\partial C_i(n_i)}{\partial n_i}}{2+\frac{(1-D(.))\acute{d(.)}}{[d(.)]^2}}$. From Assumption 5, $\frac{(1-D(.))\acute{d(.)}}{[d(.)]^2} > -1$, so that $\lim_{n_i \to \bar{n}_i} \frac{\partial W_2^*(T_g,T_s,n_i,n_e)}{\partial n_i} > 0$. From the above, one can see that $\lim_{n_i \to 0} \frac{\partial W_2^*(T_g,T_s,n_i,n_e)}{\partial n_i} > \frac{\partial C_i(n_i)}{\partial n_i}$ and $\lim_{n_i \to \bar{n}_i} \frac{\partial W_2^*(T_g,T_s,n_i,n_e)}{\partial n_i} < \lim_{n_i \to \bar{n}_i} \frac{\partial C_i(n_i)}{\partial n_i}$. □

A.2 Details on how to get Equations (2.28) and (2.29)

This simplification is a result of Envelop theorem. It is shown for the general training only. The arguments for specific training are the same. To show this note from (2.19) that $\frac{\partial f(T_g,T_s,n_i^*)}{\partial T_g} = \frac{\partial G(T_g,n_i)}{\partial T_g}|_{n_i=n_i^*} + \frac{\partial f(T_g,T_s,n_i^*)}{\partial n_i}\frac{\partial n_i^*}{\partial T_g}$. Where the first part is the direct effects of training on worker's value for a given labour supply and the second part is indirect effects that arise through changes in the labour supply. Similarly, the wage changes can be split into direct and indirect effects as $\frac{\partial(W_E=G(T_g,n_e^*))}{\partial T_g}=\frac{\partial G(T_g,n_e)}{\partial T_g}|_{n_e=n_e^*}+\frac{\partial G(T_g,n_e^*)}{\partial n_e}\frac{\partial n_e^*}{\partial T_g}$ and $\frac{\partial W_2^*}{\partial T_g}=\frac{\partial W_2^*}{\partial T_g}|_{n_i=n_i^*}+\frac{\partial W_2^*}{\partial n_i}\frac{\partial n_i^*}{\partial T_g}$. Using this, the last term in the curly brackets of (2.27) can be written as $\frac{\partial G(T_g,n_i)}{\partial T_g}|_{n_i=n_i^*}+\frac{\partial f(T_g,T_s,n_i^*)}{\partial n_i}\frac{\partial n_i^*}{\partial T_g}-\frac{\partial W_2^*}{\partial T_g}|_{n_i=n_i^*}-\frac{\partial W_2^*}{\partial n_i}\frac{\partial n_i^*}{\partial T_g}=\oplus$. But the wage change due to the labour supply in Lemma 1 can be re-written as $\frac{\partial W_2^*}{\partial n_i}=\frac{\partial f(T_g,T_s,n_i)}{\partial n_i}+[1+\frac{(1-D(.))\acute{d(.)}}{[d(.)]^2}][\frac{\partial C_i(n_i)}{\partial n_i}-\frac{\partial W_2^*}{\partial n_i}]$. From the equilibrium condition for labour supply, $\frac{\partial W_2^*}{\partial n_i}=\frac{\partial C_i(n_i)}{\partial n_i}$, this gives $\frac{\partial W_2^*}{\partial n_i}\frac{\partial n_i^*}{\partial T_g}=\frac{\partial f(T_g,T_s,n_i^*)}{\partial n_i}\frac{\partial n_i^*}{\partial T_g}$. Thus, one is left with only the direct effects of training on the output and wage; $\frac{\partial G(T_g,n_i)}{\partial T_g}|_{n_i=n_i^*}-\frac{\partial W_2^*}{\partial T_g}|_{n_i=n_i^*}=\oplus$ in the curly brackets. Similarly, the bracketed term on the left hand side of (2.27), i.e. $\frac{\partial W_2^*}{\partial T_g}-\frac{\partial G(T_g,n_e^*)}{\partial T_g}+\frac{\partial C_e(n_e^*)}{\partial T_g}-\frac{\partial C_i(n_i^*)}{\partial T_g}$ becomes $\frac{\partial W_2^*}{\partial T_g}|_{n_i=n_i^*}+\frac{\partial W_2^*}{\partial n_i}\frac{\partial n_i^*}{\partial T_g}-\frac{\partial G(T_g,n_e)}{\partial T_g}|_{n_e=n_e^*}-\frac{\partial G(T_g,n_e^*)}{\partial n_e}\frac{\partial n_e^*}{\partial T_g}+\frac{\partial C_e(n_e^*)}{\partial n_e}\frac{\partial n_e^*}{\partial T_g}-\frac{\partial C_i(n_i^*)}{\partial n_i}\frac{\partial n_i^*}{\partial T_g}=\diamond$. But from the equilibrium conditions for labour supply in the two markets, i.e. $\frac{\partial W_2^*}{\partial n_i}=\frac{\partial C_i(n_i)}{\partial n_i}$ and $\frac{\partial G(T_g,n_e)}{\partial n_e}=\frac{\partial C_e(n_e)}{\partial n_e}$, all the effects that arise indirectly through labour supply cancel out in the above expression and one is left only with the direct effects of training on the wage and outside value $\frac{\partial W_2^*}{\partial T_g}|_{n_i=n_i^*}-\frac{\partial G(T_g,n_e)}{\partial T_g}|_{n_e=n_e^*}=\diamond$. Putting all this, Equation (2.27) becomes $[1-D(.)]\{\frac{\partial G(T_g,n_i)}{\partial T_g}|_{n_i=n_i^*}-\frac{\partial W_2^*}{\partial T_g}|_{n_i=n_i^*}\}+d(.)\left(\frac{\partial W_2^*}{\partial T_g}|_{n_i=n_i^*}-\frac{\partial G(T_g,n_e)}{\partial T_g}|_{n_e=n_e^*}\right)\{f(T_s,T_g,n_i^*)-W_2^*\}=\frac{\partial C_g(T_g)}{\partial T_g}$ which is the same as when the firm ignores its training decisions impact on labour supply, i.e. labour supply is taken as given. Now putting for W_2^* from (2.22) and cancelling common terms one gets Equation (2.29).

A.3 Proof of Proposition 3

Proof. To prove Proposition 3, note that (i) follows from (2.29) as $\frac{\partial G(T_g,n_i=1)}{\partial T_g}=\frac{\partial G(T_g,n_e=1)}{\partial T_g}$. This is because of the additive separability assumption of $f(T_g,T_s,n_i)$ in T_g and T_s. (ii) follows when the condition of Proposition 2 does not hold. With $T_s=0$, equilibrium conditions for work, $\frac{\partial W_2^*}{\partial n_i}=\frac{\partial C_i(n_i^*)}{\partial n_i}$ and $\frac{\partial W_E}{\partial n_e}=\frac{\partial C_e(n_e^*)}{\partial n_e}$, imply that $n_i^*(T_g,0)=n_e^*(T_g)$ for given utility cost functions $C_i(.)\equiv C_e(.)$. This in turn implies once again that in (2.29) $\frac{\partial G(T_g,n_i^*(T_g,0))}{\partial T_g}=\frac{\partial G(T_g,n_e^*(T_g))}{\partial T_g}$. Now, when both labour supply and specific training is positive then (iii) holds, given Assumptions

3. It is because $T_s > 0$ implies $n_i^*(T_g, T_s) > n_e^*(T_g)$ as in Proposition 2. Then from the complementarity between T_g and n, $\frac{\partial G(T_g,n_i)}{\partial T_g}|_{n_i=n_i^*(T_g,T_s)} > \frac{\partial G(T_g,n_e)}{\partial T_g}|_{n_e=n_e^*(T_g)}$ must hold. Since $\frac{\partial C_g(T_g=0)}{\partial T_g} = 0$, it pays the firm to invest in general training. For (iv), the derivative of quit probability $\frac{\partial D(W_E - C_e(n_e^*(T_g)) + C_i(n_i^*(T_g,T_s)) - W_2^*)}{\partial T_g} = d(.)[\frac{\partial G(T_g,n_e)}{\partial T_g}|_{n_e=n_e^*} + \frac{\partial G(T_g,n_e^*)}{\partial n_e}\frac{\partial n_e^*}{\partial T_g} - \frac{\partial C_e(n_e^*)}{\partial n_e}\frac{\partial n_e^*}{\partial T_g} + \frac{\partial C_i(n_i^*)}{\partial n_i}\frac{\partial n_i^*}{\partial T_g} - \frac{\partial W_2^*}{\partial T_g}|_{n_i=n_i^*} - \frac{\partial W_2^*}{\partial n_i}\frac{\partial n_i^*}{\partial T_g}] = d(.)[\frac{\partial G(T_g,n_e)}{\partial T_g}|_{n_e=n_e^*} - \frac{\partial W_2^*}{\partial T_g}|_{n_i=n_i^*}]$. But $\frac{\partial G(T_g,n_e)}{\partial T_g}|_{n_e=n_e^*} < \frac{\partial W_2^*}{\partial T_g}|_{n_i=n_i^*}$ because $\frac{\partial W_2^*}{\partial T_g}|_{n_i=n_i^*} = \left[\frac{\partial G(T_g,n_i)}{\partial T_g}|_{n_i=n_i^*} + [1 + \frac{(1-D(.))\acute{d(.)}}{[d(.)]^2}]\frac{\partial G(T_g,n_e)}{\partial T_g}|_{n_e=n_e^*}\right] / \left[2 + \frac{(1-D(.))\acute{d(.)}}{[d(.)]^2}\right]$. This is a sort of weighted average and is always less than $\frac{\partial G(T_g,n_i)}{\partial T_g}|_{n_i=n_i^*}$ because $\frac{\partial G(T_g,n_i)}{\partial T_g}|_{n_i=n_i^*} > \frac{\partial G(T_g,n_e)}{\partial T_g}|_{n_e=n_e^*}$. This ensures $\frac{\partial D(.)}{\partial T_g} < 0$. Finally, to prove (v), it is enough to show that the cross partial derivative of profit function in (2.25) is positive. For this, the first order condition for general training in (2.29) can be written, after ignoring the equality sign, as $[1 - D(.)]\{\frac{\partial G(T_g,n_i)}{\partial T_g}|_{n_i=n_i^*} - \frac{\partial G(T_g,n_e)}{\partial T_g}|_{n_e=n_e^*}\} - \frac{\partial C_g(T_g)}{\partial T_g}$. Differentiating this with respect to T_s and using equilibrium conditions for work gives $d(.)\{\frac{\partial G(T_g,n_i)}{\partial T_g}|_{n_i=n_i^*} - \frac{\partial G(T_g,n_e)}{\partial T_g}|_{n_e=n_e^*}\}\frac{\partial W_2^*}{\partial T_s}|_{n_i=n_i^*} + [1-D(.)]\{\frac{\partial\left[\frac{\partial G(T_g,n_i)}{\partial T_g}|_{n_i=n_i^*}\right]}{\partial T_s}\}$. The first term of this is positive due to (iii) and the increasing wage in specific training as shown in Proposition 4. The second term is also positive given Assumption 3. □

A.4 Proof of Proposition 4

Proof. The first part is clear from Equations (2.30-2.32) as Assumption 5 ensures $\frac{(1-D(.))\acute{d(.)}}{[d(.)]^2} > -1$. Moreover, the effects of training on production $f(.)$ can be written as $\frac{\partial G(T_g,n_i)}{\partial T_g}|_{n_i=n_i^*} + \frac{\partial f(T_g,T_s,n_i^*)}{\partial n_i}\frac{\partial n_i^*}{\partial T_g}$. A similar expression can be written for W_E as is clear from the proof of Proposition 3. Using this and the equilibrium conditions for labour supply, the wage change from general training becomes $\frac{\partial W_2^*}{\partial T_g} = \frac{\frac{\partial G(T_g,n_i)}{\partial T_g}|_{n_i=n_i^*} + [1 + \frac{(1-D(.))\acute{d(.)}}{[d(.)]^2}]\frac{\partial G(T_g,n_e)}{\partial T_g}|_{n_e=n_e^*}}{2 + \frac{(1-D(.))\acute{d(.)}}{[d(.)]^2}} + \frac{\partial f(T_g,T_s,n_i^*)}{\partial n_i}\frac{\partial n_i^*}{\partial T_g}$. Similar manipulation gives the wage effects of job-specific training $\frac{\partial W_2^*}{\partial T_s} = \frac{\frac{\partial S(T_s,n_i)}{\partial T_s}|_{n_i=n_i^*}}{2 + \frac{(1-D(.))\acute{d(.)}}{[d(.)]^2}} + \frac{\partial f(T_g,T_s,n_i^*)}{\partial n_i}\frac{\partial n_i^*}{\partial T_s}$. Thus, for given productivities $\frac{\partial f(T_g,T_s,n_i)}{\partial T_g} \equiv \frac{\partial f(T_g,T_s,n_i)}{\partial T_s}$, part (ii) must hold, i.e. $\frac{\partial W_2^*}{\partial T_g} > \frac{\partial W_2^*}{\partial T_s}$. Moreover, for $\frac{(1-D(.))\acute{d(.)}}{[d(.)]^2} > -1$, the firm charges rent on both T_g and T_s as $\frac{\partial f(T_g,T_s,n_i^*)}{\partial T_g} = \frac{\partial G(T_g,n_i)}{\partial T_g}|_{n_i=n_i^*} + \frac{\partial f(T_g,T_s,n_i^*)}{\partial n_i}\frac{\partial n_i^*}{\partial T_g} > \frac{\partial W_2^*}{\partial T_g}$ and $\frac{\partial f(T_g,T_s,n_i^*)}{\partial T_s} > \frac{\partial W_2^*}{\partial T_s}$ from the above expressions and part (iii) of Proposition 3. Regarding the wage effects of labour supply, the second term in (2.32) cancels out due to the optimality condition for work. This implies that the wage change due to an hour of work is equal to its marginal contribution to the production, i.e. $\frac{\partial W_2^*}{\partial n_i} = \frac{\partial f(T_g, T_s, n_i^*)}{\partial n_i}$, in the equilibrium. □

A.5 Specific Examples

In such studies, getting closed form solution with specific functional form is difficult (Acemoglu and Pischke, 1996, Black and Loewenstein, 1997, Acemoglu and Pischke, 1998, Kessler and Lülfesmann, 2006). That is why most of these studies rely on general functional forms and just ensures existence of a solution. Acemoglu and Pischke (1996) give specific example to show multiplicity of equilibria by assuming uniform distribution and only two possible values for their endogenous variables ability and training in their example. In this model, because of the additive nature of the production function, analytical solution of the model becomes even more difficult. But just to develop an understanding of the solution two examples are given.

Example 1

This example first continues with a general $D(.)$ as it will help in understanding some of the results. Then it takes a uniform distribution to get the final two equations in two unknowns T_g and T_s. The production function in (2.19) is taken as $f(T_g, T_s, n_i) = n_i T_g + n_i T_s^\gamma$, where $\gamma \in (0, 1)$. Thus, the external wage of a worker who quits is $G(T_g, n_e) = n_e T_g$. The utility costs are assumed to be $C_i(n_i) = a n_i^2$ and $C_e(n_e) = b n_e^2$. Similarly, the training costs are assumed $C_s(T_s) = c T_s^2$ and $C_g(T_g) = d T_g^2$. Using the model solution strategy, this will give a wage of $W_2^* = n_i T_g + n_i T_s^\gamma - \frac{[1-D(.)]}{d(.)}$. Thus, it gives $\frac{\partial W_2^*}{\partial n_i} = \frac{T_g + T_s^\gamma}{2 + \frac{(1-D(.))\acute{d(.)}}{[d(.)]^2}} + \frac{[1 + \frac{(1-D(.))\acute{d(.)}}{[d(.)]^2}] 2 a n_i}{2 + \frac{(1-D(.))\acute{d(.)}}{[d(.)]^2}}$, whereas $\frac{\partial C_i(n_i)}{\partial n_i} = 2 a n_i$. Equating this ensures equilibrium labour supply of $n_i^* = \frac{T_g + T_s^\gamma}{2a}$. Similarly, in the external market one gets $n_e^* = \frac{T_g}{2b}$. The first order conditions for trainings are $[1-D(.)][\frac{T_g+T_s^{\gamma}}{2a}-\frac{T_g}{2b}]=2dT_g$ and $[1-D(.)]\frac{T_g+T_s^{\gamma}}{2a}\gamma T_s^{\gamma-1}=2cT_s$. From this, one can verify some of the results above. For example, $\frac{\partial G(T_g,n_i)}{\partial T_g}|_{n_i=n_i^*}=\frac{T_g+T_s}{2a}>\frac{T_g}{2b}=\frac{\partial G(T_g,n_e)}{\partial T_g}|_{n_e=n_e^*}$ given that $a=b$. General and specific trainings are incentive complements for the firm etc. Solving the two first order conditions, using $a=b$, gives $T_g^*=\frac{1-D(.)}{4ad}(T_s^*)^{\gamma}$ and $T_s^*=\left(\frac{\gamma(1-D(.))}{4ac}[\frac{1-D(.)}{4ad}+1]\right)^{\frac{1}{2-2\gamma}}$. From this, one can see that for a given quit probability all the cost parameters are having negative effects on both general and specific trainings. Furthermore, from $T_g^*=\frac{1-D(.)}{4ad}(T_s^*)^{\gamma}$, one can see that T_g^* increases directly due to increase in T_s^* and indirectly due to decreasing $D(.)$ in trainings, as highlighted in part (v) of Proposition 3.

For a complete solution, one needs to specify the distribution of $D(.)$. Suppose that D(.) follows uniform distribution over $\theta\in[-0.5,0.5]$. Then the period 2 profit function can be written as $\Pi_2=[1-(W_E-C_e(n_e)-W_2+C_i(n_i)+0.5)]\{f(T_g,T_s,n_i)-W_2\}$. Solving this gives optimal wage of $W_2^*=\frac{f(T_s,T_g,n_i)+W_E-C_e(n_e)+C_i(n_i)-0.5}{2}$. To confirm that the labour supply does not depend on the distribution, this wage can be used in the labour supply optimization problem to get the same optimal labour supply as in the above paragraph. Using this wage with the first-order conditions above and $a=b$, one gets $[\frac{f(T_s,T_g,n_i)-W_E+C_e(n_e)-C_i(n_i)+0.5}{2}][\frac{T_s^{\gamma}}{2a}]=2dT_g$ and $[\frac{f(T_s,T_g,n_i)-W_E+C_e(n_e)-C_i(n_i)+0.5}{2}]\frac{T_g+T_s^{\gamma}}{2a}\gamma T_s^{\gamma-1}=2cT_s$. After putting values of different functions and optimal labour supplies from above, this becomes system of two equations in T_g and T_s. The solution can be carried only through some software as the system is complicated and cannot be solved analytically. For example, if one assumes $a=b=c=0.4$, $d=0.5$ and $\gamma=0.8$, then the following two equilibriums emerge

1. $T_g^*=0.0716$, $T_s^*=0.1420$, $D(.)=0.7269$, $W_2^*=-0.1742$, $W_E=0.0064$
2. $T_g^*=0.1978$, $T_s^*=0.3540$, $D(.)=0.6368$, $W_2^*=0.1385$, $W_E=0.0489$

When all the assumptions of Section 2.4 are respected, a third equilibrium with even higher trainings and low quit probability may exist. This example also shows that why training may not take place in economies with high quit rate. Note that in equilibrium with high quit probability, the training firm optimal wage is negative. In a credit constraint economy, this will result in a zero training.

Example 2

In example 2, I consider Cobb-Douglas production technology and linear labour and training cost functions. More specifically, the production function in (2.19) is taken as $f(T_g, T_s, n_i) = n_i^{\alpha} T_g^{1-\alpha} + n_i^{\alpha} T_s^{1-\alpha}$ with $\alpha \in (0,1)$. The quitter will get $G(T_g, n_e) = n_e^{\alpha} T_g^{1-\alpha}$. The labour and training costs are assumed to be $C_i(n_i) = n_i$, $C_e(n_e) = n_e$, $C_s(T_s) = T_s$ and $C_g(T_g) = T_g$. This gives $\frac{\partial W_2^*}{\partial n_i} = \frac{\alpha n_i^{\alpha-1}[T_g^{1-\alpha}+T_s^{1-\alpha}]}{2} + \frac{1}{2}$ with a uniform distribution function. The equilibrium labour supplies become $n_i^* = (\frac{1}{\alpha})^{\frac{1}{\alpha-1}}[T_g^{1-\alpha} + T_s^{1-\alpha}]^{\frac{1}{1-\alpha}}$ and $n_e^* = (\frac{1}{\alpha})^{\frac{1}{\alpha-1}} T_g$. The first order conditions for trainings are now $[1 - D(.)][(\frac{1}{\alpha})^{\frac{\alpha}{\alpha-1}} - (\frac{1}{\alpha})^{\frac{1}{\alpha-1}}][T_g^{1-\alpha} + T_s^{1-\alpha}]^{\frac{\alpha}{1-\alpha}} T_g^{-\alpha} = 1 + [1 - D(.)][(\frac{1}{\alpha})^{\frac{\alpha}{\alpha-1}} - (\frac{1}{\alpha})^{\frac{1}{\alpha-1}}]$ for general training and $[1 - D(.)][(\frac{1}{\alpha})^{\frac{\alpha}{\alpha-1}} - (\frac{1}{\alpha})^{\frac{1}{\alpha-1}}][T_g^{1-\alpha} + T_s^{1-\alpha}]^{\frac{\alpha}{1-\alpha}} T_s^{-\alpha} = 1$ for specific training. For quit probability, one can put the relevant functions as I did in example 1. This is a well defined system of two equations in two unknowns T_g and T_s. From this, one can show numerically that all the results of the model hold true.

Table A.1: Results from the Existing Studies

	Dearden et al. (2006)			Conti (2005)			Col. & Stan. (2008)	
	RE	S. GMM	S. GMM	FE	F-GMM	S. GMM	FE	S. GMM
Production								
Training	0.70*	1.04*	0.65	0.35*	0.31	0.41*	0.03*	0.07*
	(0.17)	(0.18)	(0.41)	(0.17)	(0.21)	(0.18)	(2.44)	(3.03)
lag	-	-	-0.97	-	-	-0.35	-	-
			(0.65)			(0.28)		
K/L	0.24*	0.33*	0.32*	0.43*	0.15*	0.34*	0.45*	0.25*
	(0.02)	(0.02)	(0.06)	(0.03)	(0.07)	(0.06)	(16.63)	(7.52)
lag	-	-	-0.10	-	-	-0.13*	-	-
			(0.06)			(0.06)		
hours/L	0.20	0.52*	0.43*	0.32	-0.36	-0.29	-	-
	(0.18)	(0.06)	(0.12)	(0.36)	(0.58)	(0.40)		
lag	-	-	-0.40*	-	-	-0.73	-	-
			(0.17)			(0.46)		
lag R&D/L	1.39*	1.54*	1.14	-	-	0.04*	-	-
	(0.43)	(0.34)	(0.67)			(0.01)		
Executive/L	-	-	-	-	-	-	0.09	1.30*
							(0.62)	(6.99)
Worker/L	-	-	-	-	-	-	-0.12*	-0.43*
							(-4.16)	(-16.93)
Wage Eq.								
Training	0.34*	0.35*	0.30*	0.22	0.20	0.12	0.02*	0.04*
	(0.12)	(0.07)	(0.15)	(0.18)	(0.20)	(0.16)	(2.04)	(2.13)
lag	-	-	0.27	-	-	-0.41	-	-
			(0.29)			(0.26)		
K/L	0.05*	0.11*	0.03	0.34*	0.09	0.19*	0.42*	0.23*
	(0.01)	(0.01)	(0.04)	(0.03)	(0.06)	(0.06)	(14.88)	(6.49)
lag	-	-	0.02	-	-	-0.05	-	-
			(0.03)			(0.03)		
hours/L	0.27*	0.49*	0.48*	0.27	-0.14	-0.61	-	-
	(0.13)	(0.03)	(0.08)	(0.37)	(0.56)	(0.45)		
lag	-	-	-0.36*	-	-	-0.69	-	-
			(0.10)			(0.41)		
lag R&D/L	-0.36	0.44*	0.55	-	-	0.02*	-	-
	(0.28)	(0.18)	(0.33)			(0.01)		
Executive/L	-	-	-	-	-	-	0.01	1.02*
							(0.08)	(6.02)
Worker/L	-	-	-	-	-	-	-0.12*	-0.32*
							(-4.33)	(-13.06)
No. of obs	968	883	883	633	456	456	26312	15306

Note: Estimates for production function and wage equation. Significance at 5% level is indicated by *. Colombo and Stanca (2008) give t-statistics in parenthesis whereas in the other studies heteroskedasticity robust standard errors are reported.

Table A.2: Association between the Productivity and Wages for Entrants and Quitters

	Entrants (OLS)		Entrants (ACF)		Quitters (OLS)		Quitters (ACF)	
	Prod.	Wages	Prod.	Wages	Prod.	Wages	Prod.	Wages
University/Labour	0.165	0.101	0.106	0.081	-0.136	-0.054	-0.086	-0.038
	(0.105)	(0.109)	(0.078)	(0.094)	(0.081)	(0.087)	(0.059)	(0.076)
High school/Labour	0.026	0.012	0.020	0.011	-0.024	-0.009	-0.016	-0.007
	(0.045)	(0.044)	(0.045)	(0.040)	(0.017)	(0.018)	(0.033)	(0.021)
Secondary school/Labour	-0.012	-0.028	-0.014	-0.029	-0.007	0.013	-0.003	0.014
	(0.015)	(0.015)	(0.011)	(0.013)	(0.006)	(0.008)	(0.008)	(0.008)
Association	0.9941	0.9941	0.9961	0.9961	0.9811	0.9811	0.9660	0.9660

Note: Standard errors are estimated with 100 bootstrap iterations. Each row is estimated separately, including all the variables reported in Table 2.1 as regressors in the wage and production function, respectively.

Appendix B

Appendix to Chapter 3

B.1 Proof of Proposition 8

Proof. To prove this, define $A_i = [w^*\rho_i\phi_i\varphi + (1+g)(1-\rho_i)\phi_i - \rho_i\delta\phi_i - (1+g)^{\kappa_i}]$. Similarly, define $B_i = \left[(1-\beta)(1-\rho_i)\varphi\phi_i + (1-\rho_i)(1-\upsilon)S_i\left(\frac{(w^*+1)\varphi}{S_i\upsilon}\right)^{\frac{\upsilon}{\upsilon-1}}\phi_i\right]$ and noting that $\hat{Y}^* = \left(\frac{\beta}{w^*}\right)^{\frac{\beta}{1-\beta}}$. Hence $\hat{h}_i^*(w^*) = \frac{B_i\left(\frac{\beta}{w^*}\right)^{\frac{\beta}{1-\beta}}}{NA_i}$. Then $\frac{\partial\hat{h}_i^*(w^*)}{\partial\phi_i} = -\frac{(1+g)^{\kappa_i}}{\phi_i^2}\frac{B_i\left(\frac{\beta}{w^*}\right)^{\frac{\beta}{1-\beta}}}{N(A_i)^2} < 0$ and $\frac{\partial\hat{h}_i^*(w^*)}{\partial\sigma_i} = \rho_i(1+g)^{\kappa_i}ln(1+g)\frac{B_i\left(\frac{\beta}{w^*}\right)^{\frac{\beta}{1-\beta}}}{N(A_i)^2} > 0$ since B_i is positive. This also implies that $\frac{\partial(\frac{\hat{Y}^*}{N\hat{h}_i^*})}{\partial\phi_i} > 0$ and $\frac{\partial(\frac{\hat{Y}^*}{N\hat{h}_i^*})}{\partial\sigma_i} < 0$ as $\frac{\hat{Y}^*(w^*)}{N\hat{h}_i^*(w^*)} = \frac{A_i}{B_i}$. Then from (3.23), one gets $\frac{\partial m_i^*(w^*)}{\partial\phi_i} < 0$ and $\frac{\partial m_i^*(w^*)}{\partial\sigma_i} > 0$. $\frac{\partial\hat{h}_i^*(w^*)}{\partial\phi_i} < 0$ and $\frac{\partial m_i^*(w^*)}{\partial\phi_i} < 0$ together imply $\frac{\partial\hat{I}_i^*(w^*)}{\partial\phi_i} < 0$. Similarly, $\frac{\partial\hat{h}_i^*(w^*)}{\partial\sigma_i} > 0$ and $\frac{\partial m_i^*(w^*)}{\partial\sigma_i} > 0$ together imply $\frac{\partial\hat{I}_i^*(w^*)}{\partial\sigma_i} > 0$.

From (3.23), $\frac{\partial z_i^*(w^*)}{\partial\phi_i} = \frac{\rho_i(1+g)^{\kappa_i}}{\phi_i^2(1-\rho_i)\varphi} > 0$ and $\frac{\partial z_i^*(w^*)}{\partial\sigma_i} = -\frac{\rho_i^2(1+g)^{\kappa_i}ln(1+g)}{\phi_i(1-\rho_i)\varphi} < 0$. This then implies that $\frac{\partial l_i^*(w^*)}{\partial\phi_i} > 0$ and $\frac{\partial l_i^*(w^*)}{\partial\sigma_i} < 0$ from (3.22). $\frac{\partial T_i^*(w^*)}{\partial\phi_i} > 0$ and $\frac{\partial T_i^*(w^*)}{\partial\sigma_i} < 0$ are clear from (3.23). The results about $l_i^*(w^*)$ and $T_i^*(w^*)$ together imply $\frac{\partial x_i^*(w^*)}{\partial\phi_i} < 0$ and $\frac{\partial x_i^*(w^*)}{\partial\sigma_i} > 0$. This is also clear from differentiation of $x_i^*(w^*)$ in (3.23). Finally, decreasing working hours and human capital together imply decreasing gross total wage in the patience level ϕ_i and inter-temporal elasticity of substitution $1/\sigma_i$. □

B.2 Proof of Proposition 9

Proof. First, as is clear from (3.23), $z_i^*(w^*)$ is independent of S_i. Then from (3.22), this implies that $l_i^*(w^*)$ is independent of S_i on the balanced growth path. For part (ii), using the notations as in Proposition 8, one can get $\frac{\partial \hat{h}_i^*(w^*)}{\partial S_i} = \frac{\frac{1}{1-\upsilon} S_i^{\frac{\upsilon}{1-\upsilon}} \Lambda_i \left(\frac{\beta}{w^*}\right)^{\frac{\beta}{1-\beta}}}{N[w^* \rho_i \phi_i \varphi + (1+g)(1-\rho_i)\phi_i - \rho_i \delta \phi_i - (1+g)^{\kappa_i}]} > 0$ for all $w^* \geq \hat{w}^*$. Here $\Lambda_i = [(1-\rho_i)(1-\upsilon)\left(\frac{(w^*+1)\varphi}{\upsilon}\right)^{\frac{\upsilon}{\upsilon-1}} \phi_i]$ and is positive for all $w^* \geq 0$. Since $z_i^*(w^*) = \frac{\hat{c}_i^*(w^*)}{\hat{h}_i^*(w^*)}$ and is independent of S_i. This along with $\frac{\partial \hat{h}_i^*(w^*)}{\partial S_i} > 0$ implies that $\frac{\partial \hat{c}_i^*(w^*)}{\partial S_i} > 0$. $\frac{\partial \hat{h}_i^*(w^*)}{\partial S_i} > 0$ also implies that $\frac{\partial (\frac{\hat{Y}^*}{N\hat{h}_i^*})}{\partial S_i} < 0$. Then, the first derivative of $T_i^*(w^*)$ becomes $\frac{\partial T_i^*(w^*)}{\partial S_i} = \left(\frac{(w^*+1)\varphi}{\upsilon}\right)^{\frac{1}{\upsilon-1}} [\frac{1}{1-\upsilon} S_i^{\frac{\upsilon}{1-\upsilon}} \frac{\hat{Y}^*}{N\hat{h}_i^*} + S_i^{\frac{1}{1-\upsilon}} \frac{\partial (\frac{\hat{Y}^*}{N\hat{h}_i^*})}{\partial S_i}]$. Here the first term in the sum is positive whereas the second is negative due to $\frac{\partial (\frac{\hat{Y}^*}{N\hat{h}_i^*})}{\partial S_i} < 0$. But the absolute value of the negative term can be shown to be less than the positive term by contradiction. Suppose to the contrary that $\frac{1}{1-\upsilon} S_i^{\frac{\upsilon}{1-\upsilon}} \frac{\hat{Y}^*}{N\hat{h}_i^*} < -S_i^{\frac{1}{1-\upsilon}} \frac{\partial (\frac{\hat{Y}^*}{N\hat{h}_i^*})}{\partial S_i}$. Since $\frac{\hat{Y}^*}{N\hat{h}_i^*} = \frac{A_i}{B_i}$. Then for a given $w^* \geq \hat{w}^*$, putting values and cancelling common positive terms in the above inequality one can get $S_i^{\frac{\upsilon}{1-\upsilon}} < S_i^{\frac{1+\upsilon}{1-\upsilon}} \frac{\Lambda_i}{B_i}$. Taking $S_i^{\frac{1}{1-\upsilon}}$ common from B_i one gets $1 < \frac{\Lambda_i}{C_i + \Lambda_i}$, where $C_i = (1-\beta)(1-\rho_i)\varphi \phi_i / S_i^{\frac{1}{1-\upsilon}} > 0$. But for positive Λ_i value $1 < \frac{\Lambda_i}{C_i + \Lambda_i}$ is not possible. Hence $\frac{1}{1-\upsilon} S_i^{\frac{\upsilon}{1-\upsilon}} \frac{\hat{Y}^*}{N\hat{h}_i^*} > -S_i^{\frac{1}{1-\upsilon}} \frac{\partial (\frac{\hat{Y}^*}{N\hat{h}_i^*})}{\partial S_i}$ which implies that $\frac{\partial T_i^*(w^*)}{\partial S_i} > 0$. Similarly, the derivative of $x_i^*(w^*)$ becomes $\frac{\partial x_i^*(w^*)}{\partial S_i} = [\frac{1}{1-\upsilon} S_i^{\frac{\upsilon}{1-\upsilon}} \frac{\hat{Y}^*}{N\hat{h}_i^*} + S_i^{\frac{1}{1-\upsilon}} \frac{\partial (\frac{\hat{Y}^*}{N\hat{h}_i^*})}{\partial S_i}][\varphi \left(\frac{(w^*+1)\varphi}{\upsilon}\right)^{\frac{1}{\upsilon-1}} - \upsilon \left(\frac{(w^*+1)\varphi}{\upsilon}\right)^{\frac{\upsilon}{\upsilon-1}}]/w^*\varphi < 0$. This is due to the fact that the first term is positive for $w^* \geq \hat{w}^*$ and the second term is negative as shown in Proposition 7. □

B.3 Model with Exogenous Labour Supply

If one exogenises labour supply in the model in Section 3.3, the result can change significantly. Though, when labour supply is exogenous and assumed equal to one, one cannot examine the effect of schooling differences on training time, but can analyse its effects on good investment in human capital. With exogenous labour supply, the production and profit functions of the firm become as:

$$Y_t = A_t^{1-\beta} H_t^{\beta}, \tag{B.1}$$

$$\Pi_t = \underset{H_t}{max}[A_t^{1-\beta} H_t^{\beta} - w_t H_t]. \tag{B.2}$$

The infinitely lived consumer i now maximises the following utility function in this model economy

$$maxU_i = \sum_{t=0}^{\infty} \phi^t \frac{[c_{it}]^{1-\sigma}}{1-\sigma}, \tag{B.3}$$

subject to

$$c_{it} = w_t h_{it} - I_{it} + \frac{1}{N}\Pi_t, \tag{B.4}$$

$$h_{it+1} = h_{it}^{\upsilon} \left(\frac{Y_t}{N}\right)^{1-\upsilon} + S_i I_{it} - \delta h_{it}, \tag{B.5}$$

$$I_{it}, c_{it} \geq 0, \tag{B.6}$$

and the initial conditions $h_{i0} > 0$. Note that multiplying S_i with the first instead of the second term in (B.5) will give similar results. With this change, the balanced growth equations system becomes

$$\begin{aligned}
\hat{c}_i &= \left(\beta(\sum_{i=1}^{N}\hat{h}_i)^{\beta-1}\right)\hat{h}_i - \hat{I}_i + \frac{1}{N}\left((1-\beta)(\sum_{i=1}^{N}\hat{h}_i)^{\beta}\right), \\
(1+g)^{\sigma} &= \phi S_i \left(\beta(\sum_{i=1}^{N}\hat{h}_i)^{\beta-1}\right) + \phi\upsilon\hat{h}_i^{\upsilon-1}\left(\frac{1}{N}(\sum_{i=1}^{N}\hat{h}_i)^{\beta}\right)^{1-\upsilon} - \phi\delta, \\
(1+g+\delta)\hat{h}_i &= S_i\hat{I}_i + \hat{h}_i^{\upsilon}\left(\frac{1}{N}(\sum_{i=1}^{N}\hat{h}_i)^{\beta}\right)^{1-\upsilon},
\end{aligned} \tag{B.7}$$

where the terms in the brackets are per unit wage and profit of the firm.

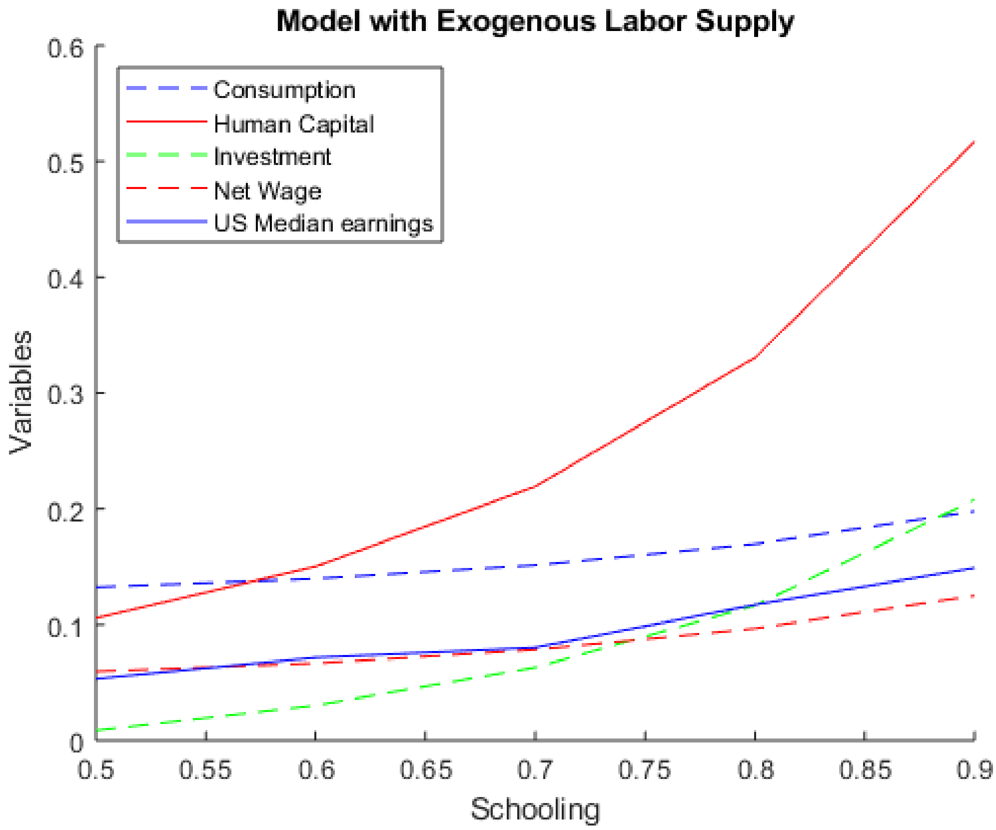

Figure B.1: Additive Separable Human Capital Function

For the numerical solution, parameter values are borrowed from Section 3.3's baseline estimates. Only, I set the value of υ equal to 0.8 implying the share of firm effects of 0.2. The results for this exercise are reported in Figure B.1. As is clear from the figure, the earlier results still hold. That is, consumption, human capital and net of investment costs wages are still increasing in the schooling S_i. The difference from earlier results is that all the functions are now steeper as compared to the endogenous labour supply specification. For example, now the share of the highest schooling individual is 39.0 percent in human capital and 25.0 in consumption. The wage function is also steeper as compared to earlier. Thus, the inequality under the exogenous labour supply is more as compared to the wage inequality under endogenous labour supply. When compared to the US's weekly earnings by education, one can see that the model generated wage distribution is slightly flatter as compared to the data on average. Another difference is that now, consumption, human capital investment and human capital can increase in pre-job schooling through preferences effects if one assumes $\sigma_i(S_i)$ and $\phi_i(S_i)$.

B.4 Complementary Human Capital and Investment

This section assumes that in future human capital generation, current human capital and goods investment are complementary. It changes the human capital production equation as $h_{it+1} = S_i I_{it}^{\varsigma} h_{it}^{\upsilon} \left(\frac{Y_t}{N}\right)^{1-\varsigma-\upsilon} - \delta h_{it}$, where $\upsilon + \varsigma \in (0,1)$. Here ς is the share of individual i investment in capital production and υ is the share of current human capital. With this change, the equations system (B.7) now becomes

$$
\begin{aligned}
\hat{c}_i &= \left(\beta(\sum_{i=1}^{N}\hat{h}_i)^{\beta-1}\right)\hat{h}_i - \hat{I}_i + \frac{1}{N}\left((1-\beta)(\sum_{i=1}^{N}\hat{h}_i)^{\beta}\right), \\
(1+g)^{\sigma} &= \phi\varsigma S_i\left(\beta(\sum_{i=1}^{N}\hat{h}_i)^{\beta-1}\right)\hat{I}_i^{\varsigma-1}\hat{h}_i^{\upsilon}\left(\frac{1}{N}(\sum_{i=1}^{N}\hat{h}_i)^{\beta}\right)^{1-\varsigma-\upsilon} + \phi\upsilon S_i\hat{I}_i^{\varsigma}\hat{h}_i^{\upsilon-1}\left(\frac{1}{N}(\sum_{i=1}^{N}\hat{h}_i)^{\beta}\right)^{1-\varsigma-\upsilon} - \phi\delta, \\
(1+g+\delta)\hat{h}_i &= S_i\hat{I}_i^{\varsigma}\hat{h}_i^{\upsilon}\left(\frac{1}{N}(\sum_{i=1}^{N}\hat{h}_i)^{\beta}\right)^{1-\varsigma-\upsilon}.
\end{aligned}
\tag{B.8}
$$

For the numerical solution, all earlier parameter values are used except υ and ς. Since the minimum value of firm or spill-over effects in the literature is 0.2 (Brown and Medoff, 1989, Abowd et al., 1999, Oi and Idson, 1999, Card et al., 2013, Barth et al., 2016, Aloi and Tournemaine, 2013), the values of υ and ς are kept equal to 0.4 each. Once again it takes five discrete values of schooling with uniform distribution, as $S_i = \{0.5, 0.6, 0.7, 0.8, 0.9\}$. The results are reported in Figure B.2. Once again, consumption, investment, human capital and net of investment costs wages are increasing at an increasing rate in schooling S_i. Compared to the results in previous specification, now all functions are even more steeper. For example, now the human capital share of highest schooling individual is approximately 49.0 percent versus the share of lowest schooling individual of 2.6 percent. Similarly, the consumption share of highest schooling individual is approximately 34.0 percent

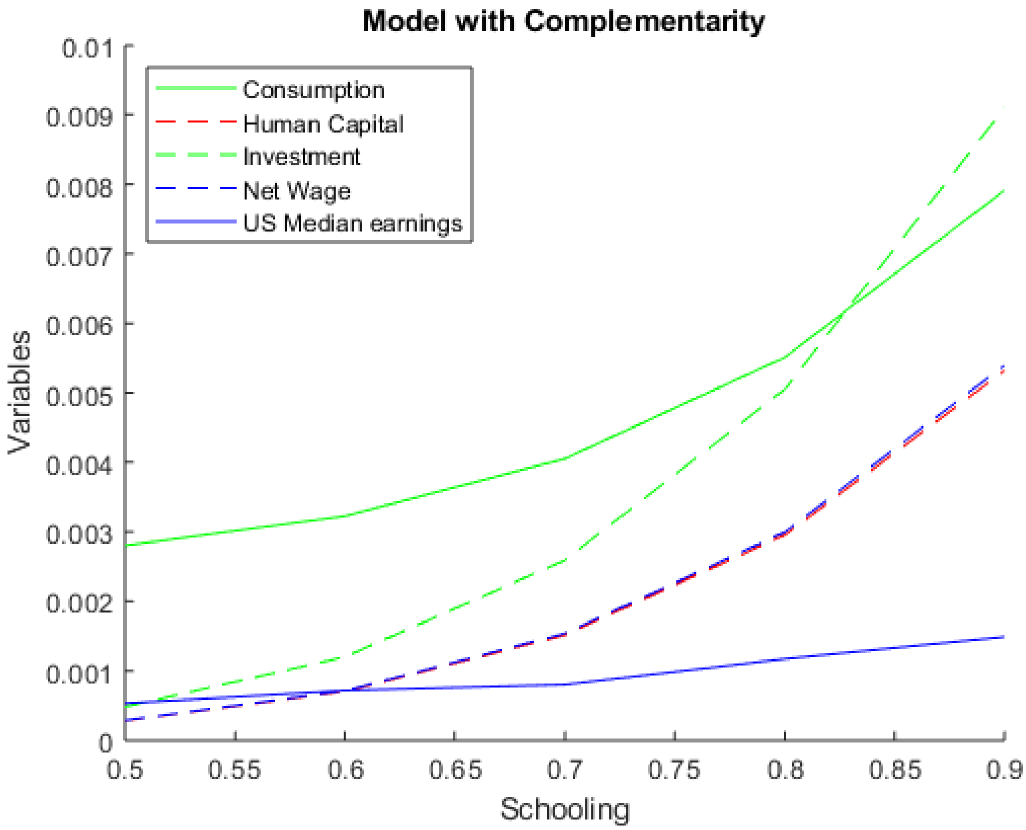

Figure B.2: Complementarity in Human Capital Function

whereas lowest schooling individual's share is 12.0 percent. The reason for this, besides the complementarity, is that now schooling directly helps current human capital in future human capital accumulation. Now the net wage function is strongly convex and steeper than the US's median earning distribution by schooling. But the weak point of this section is exogeneous labour supply and the assumption that workers only invest goods in human capital production. This exercise shows that the results can be quite different under exogenous versus endogenous labour supply assumption.

B.5 Graphs for a Baseline Model with Endogenous Labour Supply

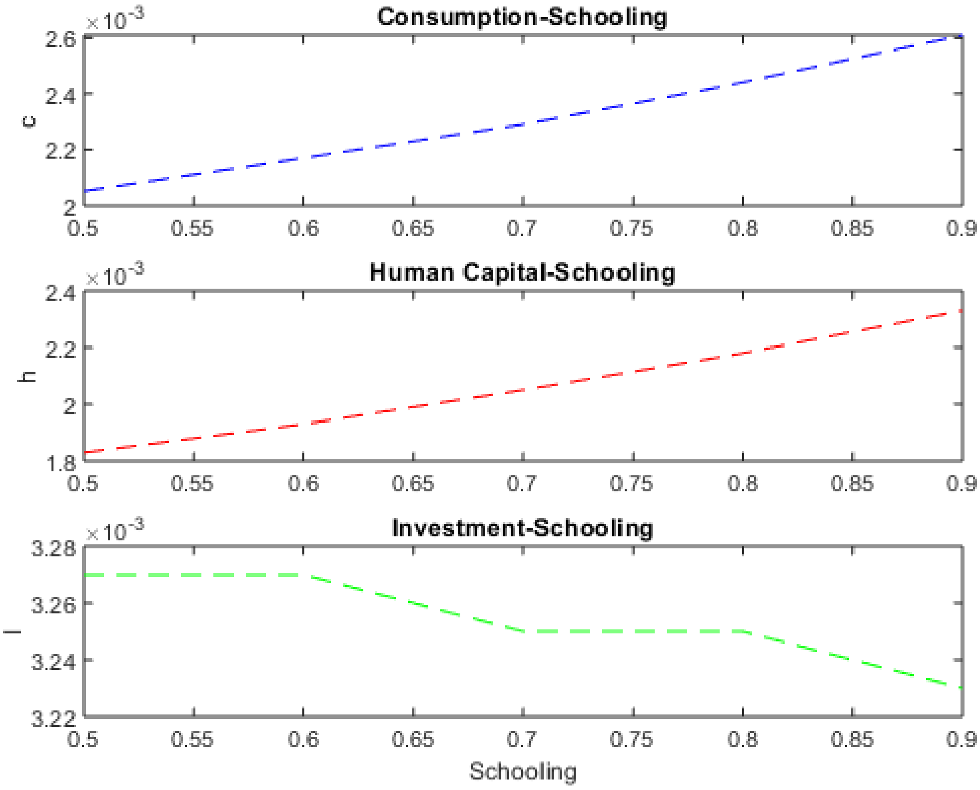

Figure B.3: Pre-Job Schooling Effects on Consumption and Skills Investment

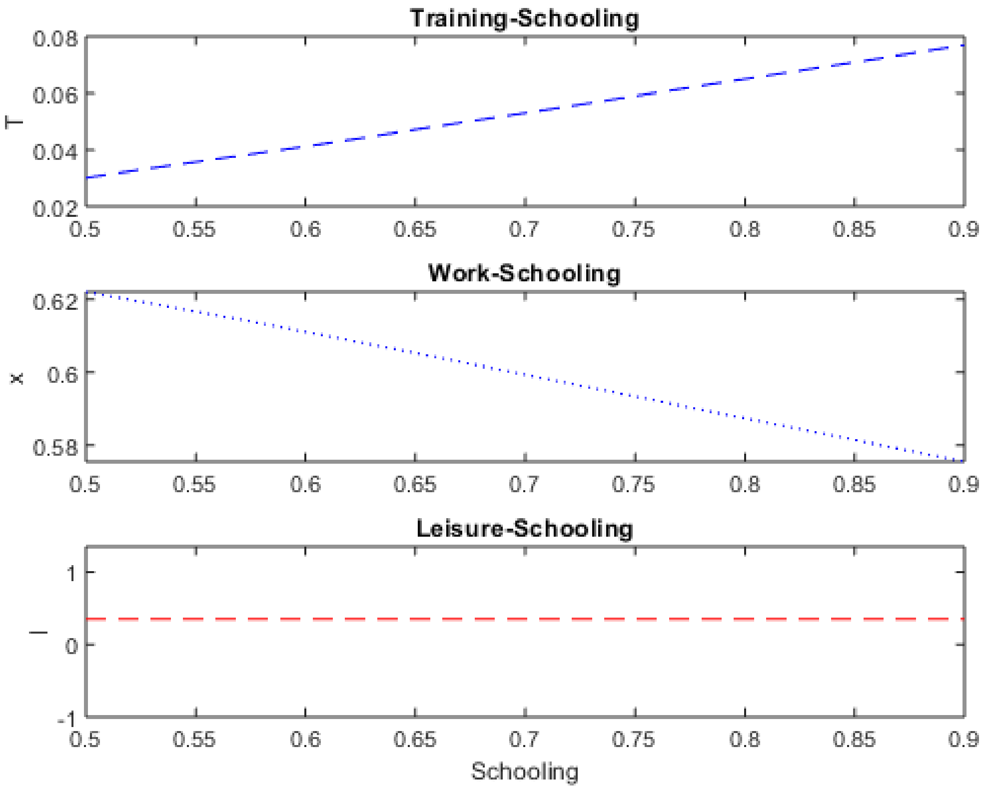

Figure B.4: Pre-Job Schooling Effects on Time Allocation

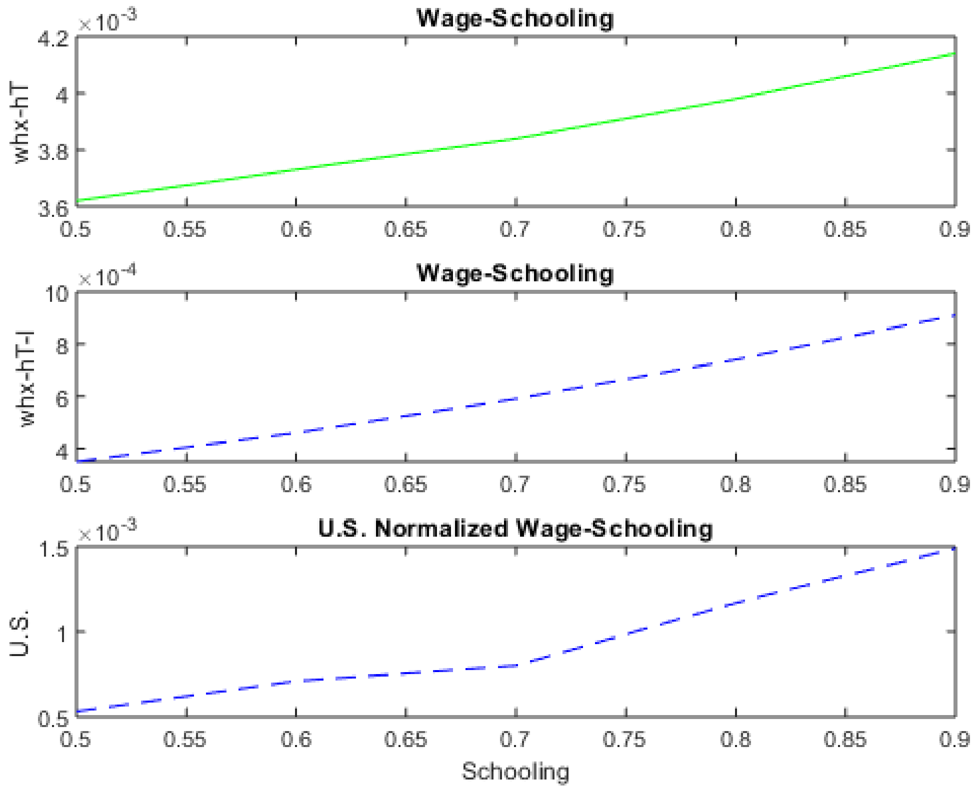

Figure B.5: Pre-Job Schooling Effects on Wages

B.6 Graphs for a Model with η in the Firm Effects

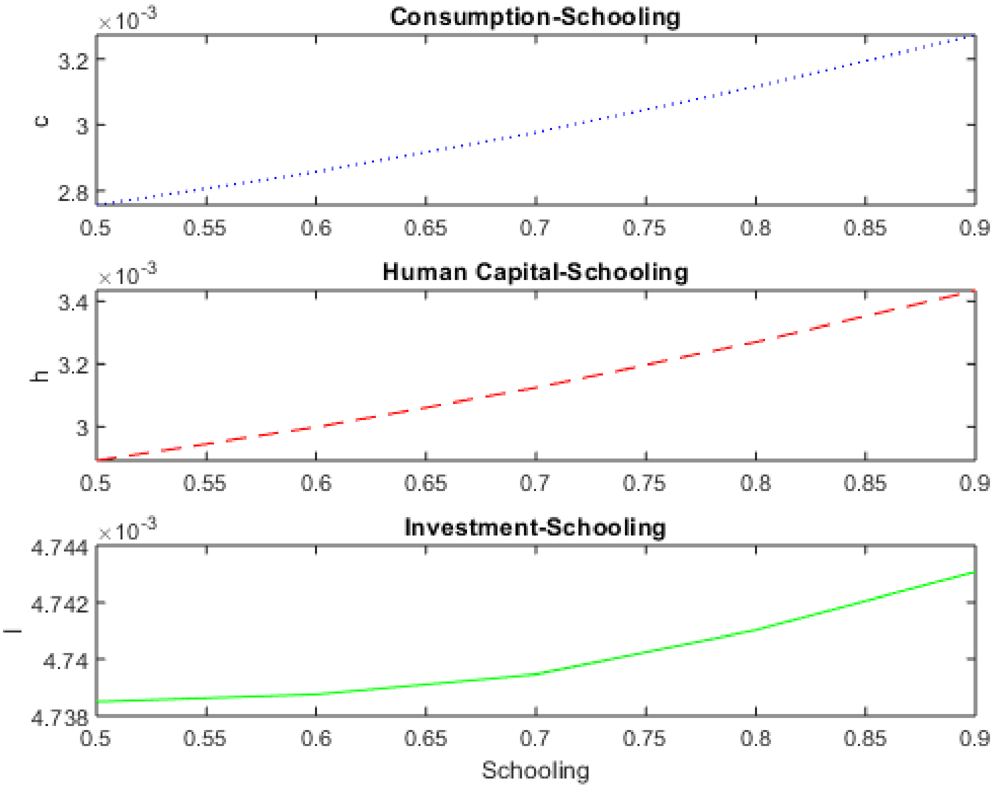

Figure B.6: Effects of a Change in Firm Effects on Consumption and Skills Investment

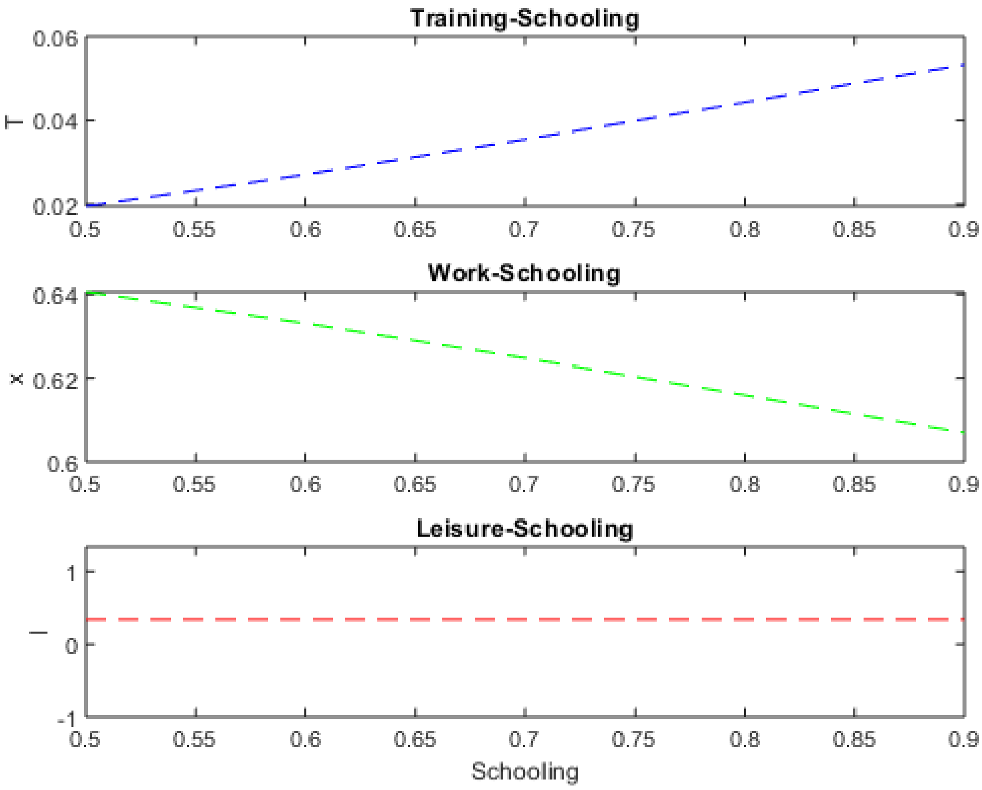

Figure B.7: Effects of a Change in Firm Effects on Time Allocation

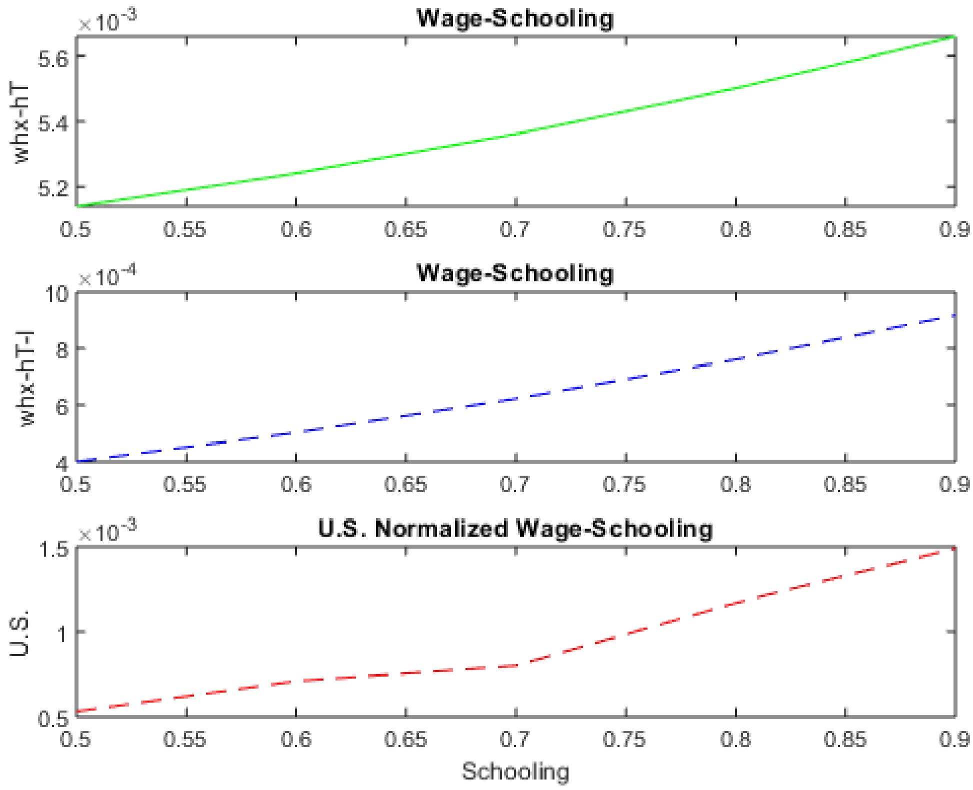

Figure B.8: Effects of a Change in Firm Effects on Wages

B.7 Graphs for a Model with an Increase in υ Value

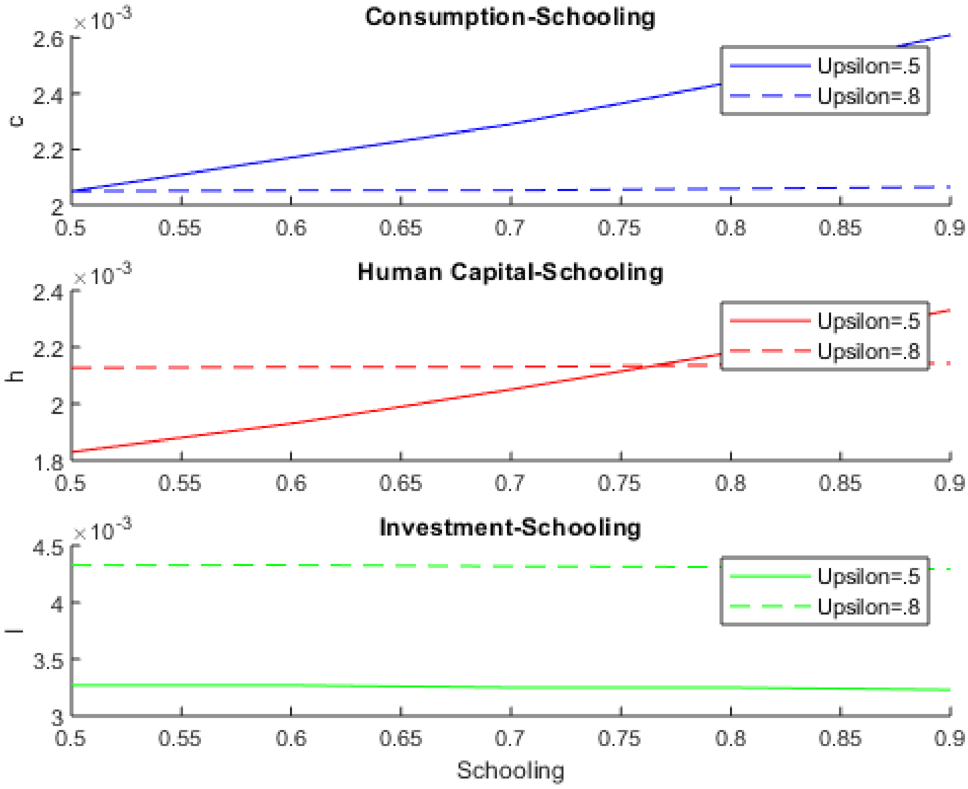

Figure B.9: Effects of a Change in υ on Consumption and Skills Investment

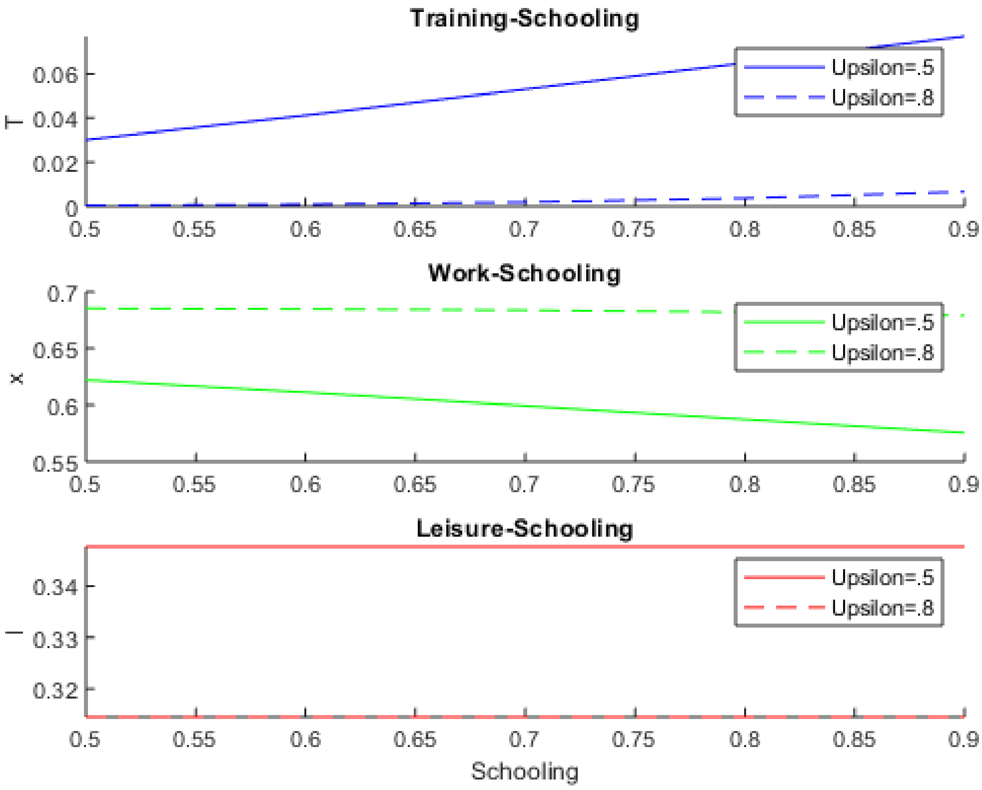

Figure B.10: Effects of a Change in υ on Time Allocation

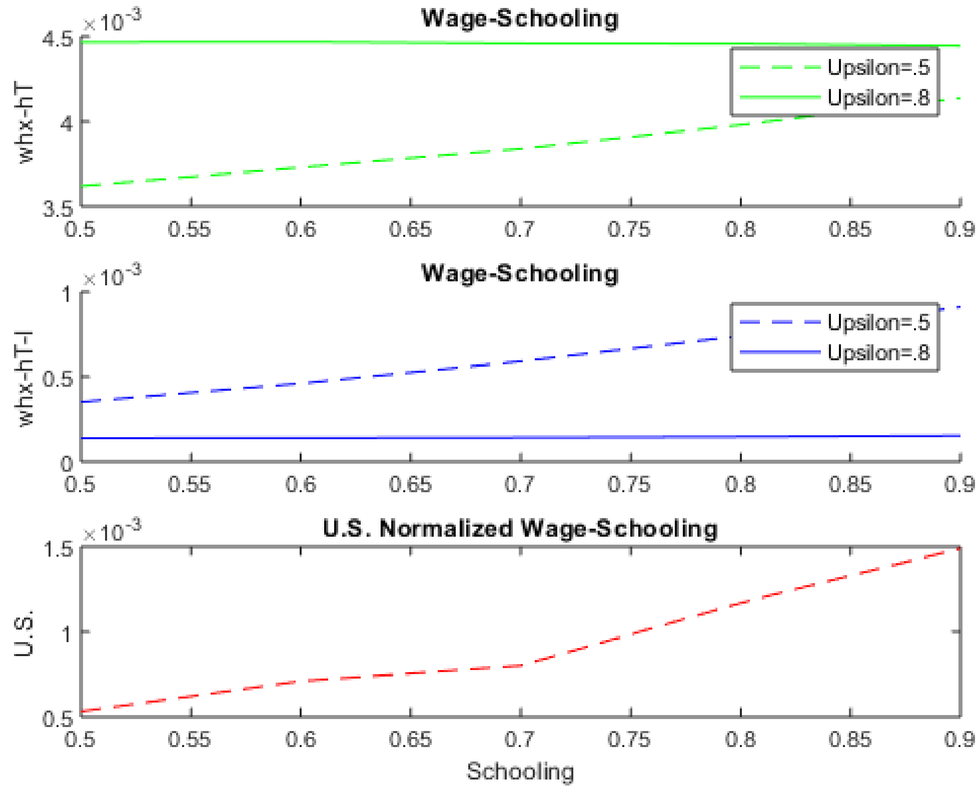

Figure B.11: Effects of a Change in υ on Wages

B.8 Graphs for a Model with Patience Effects of Schooling

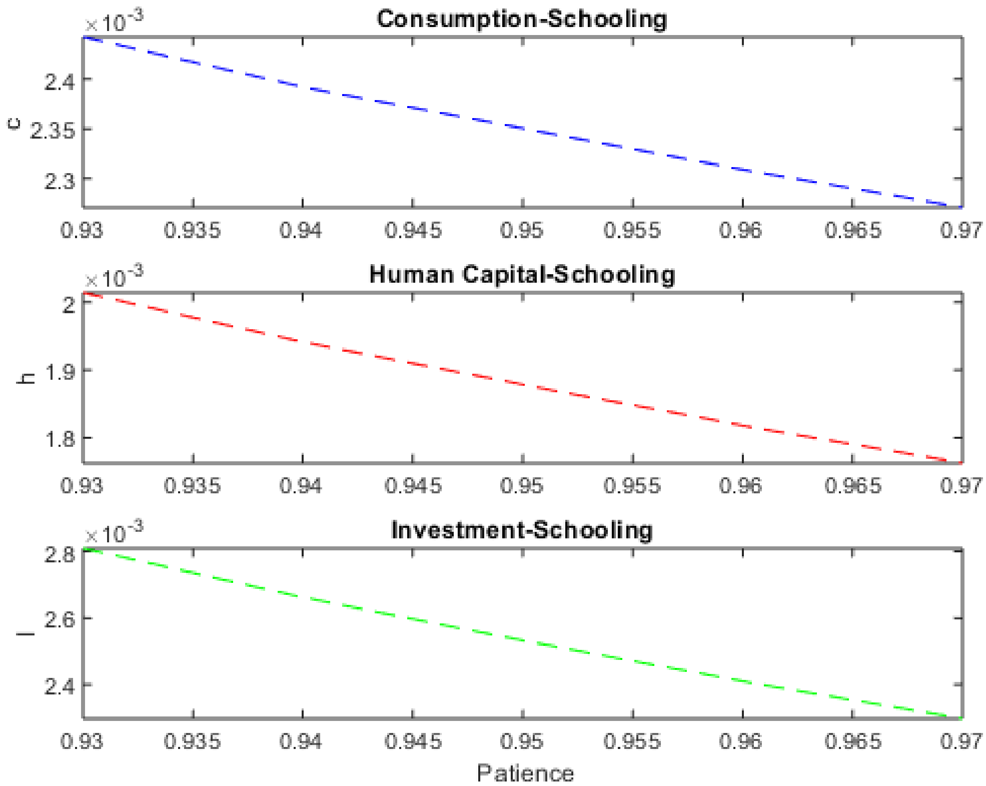

Figure B.12: Effects of Differences in $\phi(S_i)$ on Consumption and Skills Investment

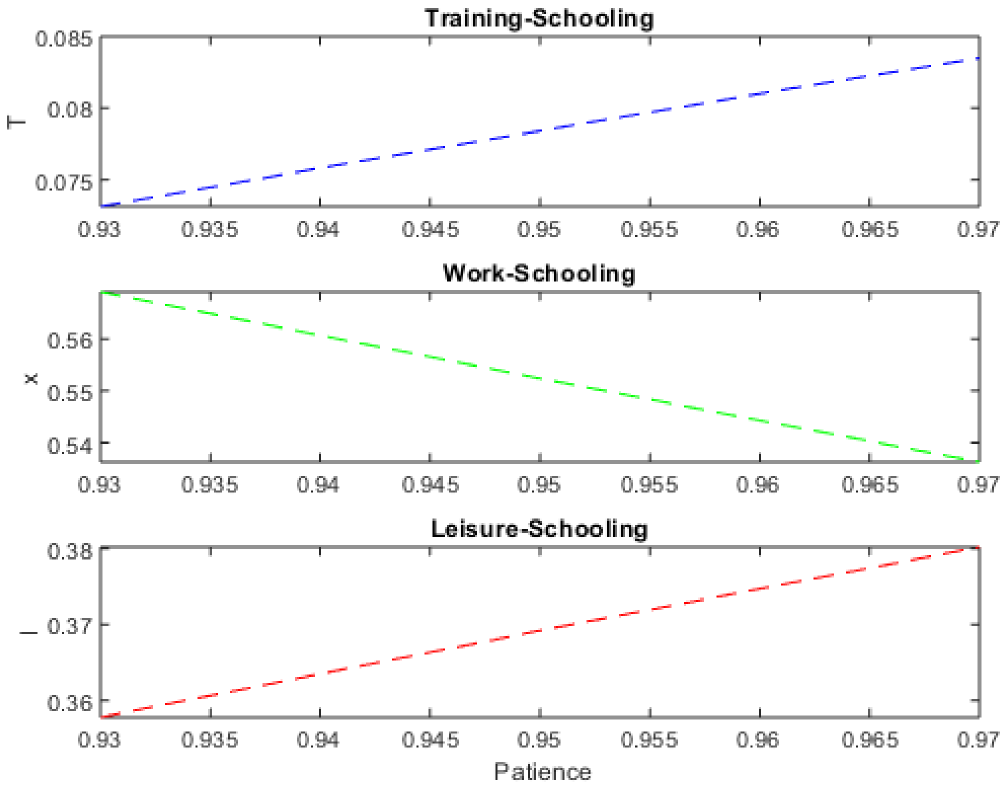

Figure B.13: Effects of Differences in $\phi(S_i)$ on Time Allocation

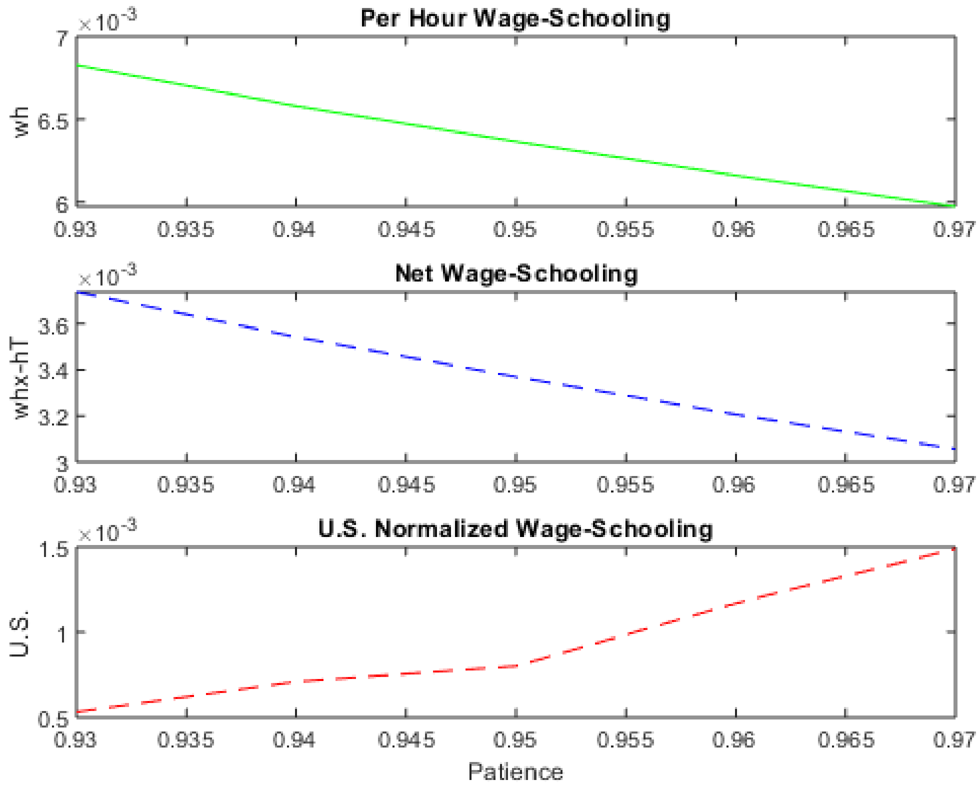

Figure B.14: Effects of Differences in $\phi(S_i)$ on Wages

B.9 Graphs for a Model with ρ Effects of Schooling

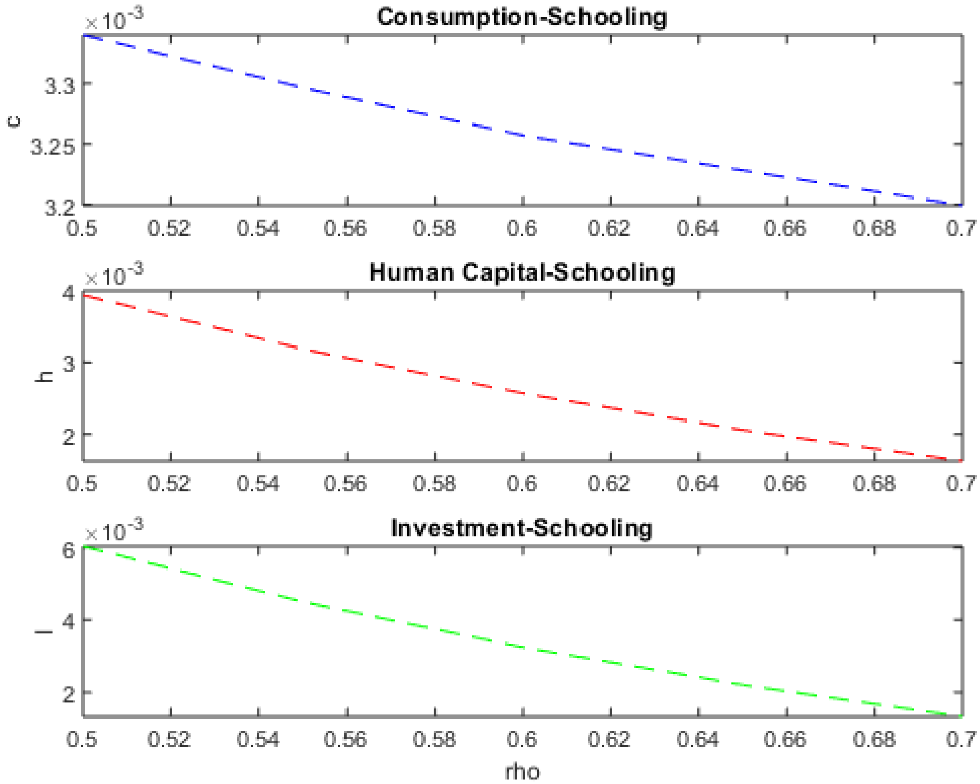

Figure B.15: Effects of Differences in $\rho(S_i)$ on Consumption and Skills Investment

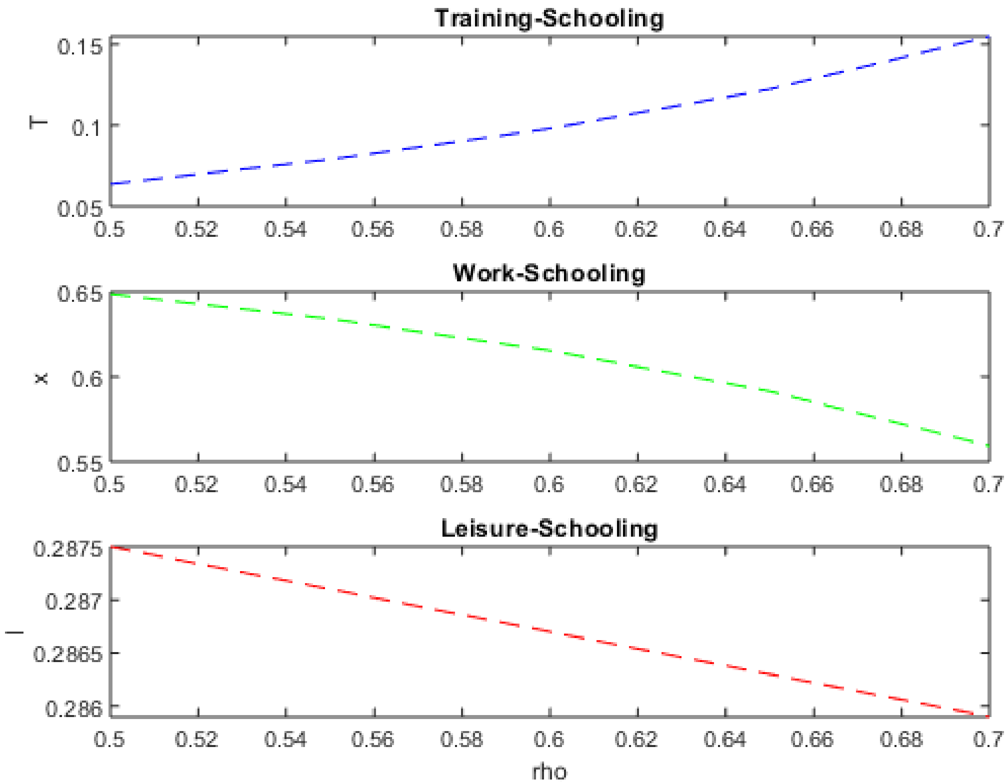

Figure B.16: Effects of Differences in $\rho(S_i)$ on Time Allocation

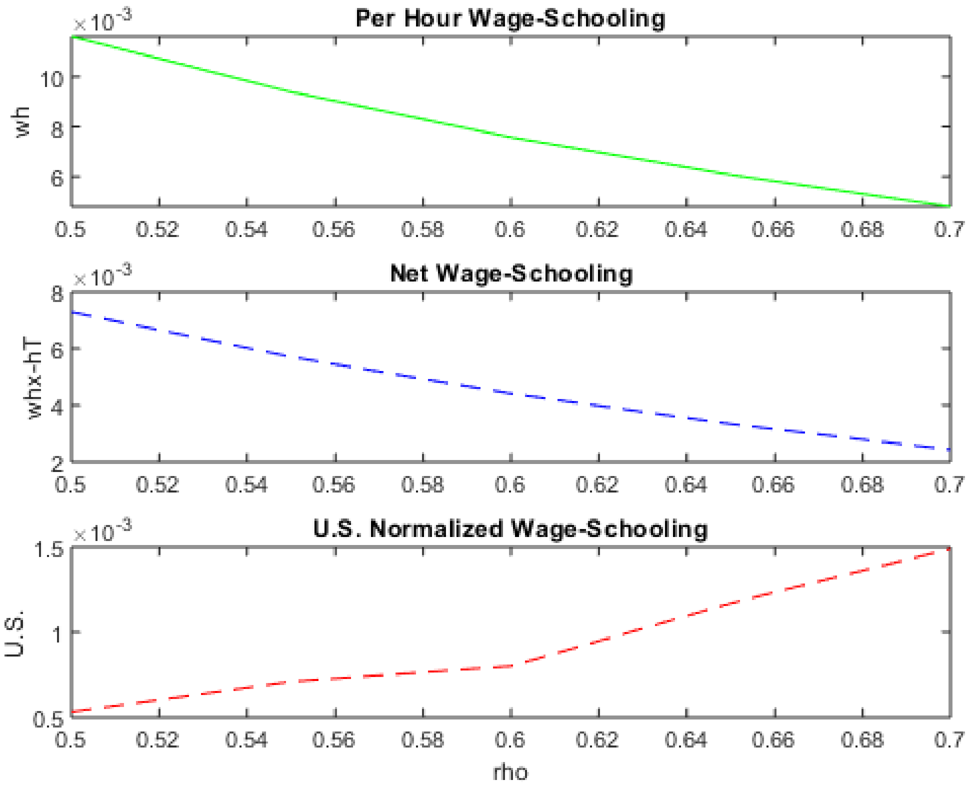

Figure B.17: Effects of Differences in $\rho(S_i)$ on Wages

Appendix C

Appendix to Chapter 4

C.1 Descriptive Statistics

Table C.1: Definition and means of variables

Variable	Definition	Mean
Train Incidence2:	No. of training courses taken over the sample period including all types of trainings	
BHPS		**0.60**
Trained Sample in BHPS		**0.90**
UKHLS		**0.80**
Trained Sample in UKHLS		**1.21**
Empl. Train Inc.2:	Number of employer provided training courses taken over the sample period	
UKHLS		**0.51**
Trained Sample in UKHLS		**0.78**
Workplace Train2:	Number of trainings taken at employment place over the sample period	
BHPS		**0.28**

Table C.1: Definition and means of variables

Variable	Definition	Mean
Trained Sample in BHPS		**0.42**
Train Incidence:	Cumulative sum of training courses taken over the sample period including all types of trainings	
BHPS		**2.20**
Trained Sample in BHPS		**3.30**
UKHLS		**2.54**
Trained Sample in UKHLS		**3.84**
Empl. Train Inc.:	Cumulative sum of employer provided trainings taken over the sample period	
UKHLS		**1.64**
Trained Sample in UKHLS		**2.48**
Workplace Train:	Cumulative sum of trainings taken at employment place over the sample period	
BHPS		**1.01**
Trained Sample in BHPS		**1.52**
Train Intensity2:	No. of training hours taken over the sample period including all types of trainings	
BHPS		**40.92**
Trained Sample in BHPS		**61.40**
UKHLS		**18.45**
Trained Sample in UKHLS		**27.92**
Monthly Wage:	Log of monthly usual gross wage	
BHPS		**7.15**
Trained Sample in BHPS		**7.24**

Table C.1: Definition and means of variables

Variable	Definition	Mean
UKHLS		**7.35**
Trained Sample in UKHLS		**7.45**
Qualification:	Nine qualifications categories with 1 denoting degree or above and 9 denotes no qualifications	
BHPS		**3.51**
Trained Sample in BHPS		**3.14**
UKHLS		**2.81**
Trained Sample in UKHLS		**2.54**
Firm Size:	Nine categories; 1 denoting firms with 1-2 workers and 9 denotes firms with 1000 and above workers	
BHPS		**4.90**
Trained Sample in BHPS		**5.11**
UKHLS		**5.10**
Trained Sample in UKHLS		**5.30**
Job Tenure:	job tenure in current firm in years	
BHPS		**6.91**
Trained Sample in BHPS		**6.45**
Other Income:	Annual non-labour income	
BHPS		**1568.39**
Trained Sample in BHPS		**1480.78**
Benefits:	Annual income from benefits and other sources	
UKHLS		**209.12**

Table C.1: Definition and means of variables

Variable	**Definition**	**Mean**
Trained Sample in UKHLS		**198.45**
Job Group:	Current job socio-economic group, ranging from large scale manager (2) to agriculture worker (18)	
UKHLS		**8.54**
Trained Sample in UKHLS		**8.28**
Industrial class:	Job two digits industrial classification (CNEF)	
UKHLS		**23.48**
Trained Sample in UKHLS		**24.17**
Unemployment:	No. of unemployment spells since last interview	
BHPS		**0.04**
Trained Sample in BHPS		**0.04**
UKHLS		**0.03**
Trained Sample in UKHLS		**0.02**
Children:	Number of children aged under 16 individual is responsible for	
UKHLS		**0.34**
Trained Sample in UKHLS		**0.36**
Second Job:	Equal one if currently holds a second job	
BHPS		**0.08**
Trained Sample in BHPS		**0.09**
UKHLS		**0.07**
Trained Sample in UKHLS		**0.08**

Table C.1: Definition and means of variables

Variable	Definition	Mean
Permanent:	Equal one if the current job is permanent	
BHPS		**0.96**
Trained Sample in BHPS		**0.96**
UKHLS		**0.94**
Trained Sample in UKHLS		**0.95**
Manager:	Equal one if the person is manager, foreman or supervisor at current job, zero otherwise	
UKHLS		**0.38**
Trained Sample in UKHLS		**0.42**
Satisfaction:	Satisfaction of job with 1 denotes least and 7 denotes most satisfied	
BHPS		**5.40**
Trained Sample in BHPS		**5.39**
UKHLS		**5.28**
Trained Sample in UKHLS		**5.27**
Age:	Age in years	
BHPS		**39.39**
Trained Sample in BHPS		**38.72**
UKHLS		**42.01**
Trained Sample in UKHLS		**41.96**

C.2 Tables for the Estimations

Table C.2: Two-Way Fixed-Effects Results from the UKHLS Panel

Variables	*OLS*	*FE*	*FE*	*FE*	*FE*	*FE*
Train Intensity	0.0001	0.0000	0.0000	-	0.0000	-0.0000
	(0.0000)*	(0.0000)	(0.0000)		(0.0000)	(0.0000)
Train Incidence	0.0123	0.0043	0.0038	-	-	0.0044
	(0.0004)*	(0.0008)*	(0.0008)*			(0.0009)*
Empl. Train Inc.	-	-	-	-	0.0103	-
					(0.0016)*	
Train Intensity2	-	-	-	-0.0000	-	-
				(0.0000)**		
Train Incidence2	-	-	-	0.0030	-	-
				(0.0007)*		
Age	0.1225	0.0555	0.0420	0.0426	0.0418	-0.9919
	(0.0010)*	(0.0172)*	(0.0120)**	(0.0196)**	(0.0196)**	(0.0234)*
Age Square	-0.0014	-0.0006	-0.0004	-0.0005	-0.0004	0.0113
	(0.0000)*	(0.0002)*	(0.0002)**	(.0002)**	(0.0002)**	(0.0003)*
Qualification	-0.3702	-0.1374	-0.0282	-0.0319	-0.0280	-0.0083
	(0.0116)*	(0.0703)**	(0.0237)	(0.0236)	(0.0237)	(0.0234)
Qualif. Square	0.0455	0.0335	0.0031	0.0034	0.0032	0.0018
	(0.0033)*	(0.0181)	(0.0024)	(0.0024)	(0.0024)	(0.0023)
Qualif. Cube	-0.0021	-0.0021	-	-	-	-
	(0.0002)*	(0.0013)				
Children	-0.2805	-0.0555	-0.0517	-0.0525	-0.0515	-0.0490
	(0.0027)*	(0.0068)*	(0.0073)*	(0.0073)*	(0.0072)*	(0.0071)*
Firm Size	-	-	0.0475	0.0477	0.0476	0.06315
			(0.0093)*	(0.0092)*	(0.0093)*	(0.010)*
F. Size Square	-	-	-0.0023	-0.0023	-0.0023	-0.0044
			(0.0008)*	(0.0008)*	(0.0008)*	(0.0010)*
Permanent	-	-	0.0709	0.0708	0.0709	0.0576
			(0.0106)*	(0.0106)*	(0.0106)*	(0.0107)*
Second Job	-	-	-0.0629	-0.0631	-0.063	-0.0519
			(0.0080)*	(0.0080)*	(0.0080)*	(0.0078)*
Satisfaction	-	-	0.0078	0.0078	0.0078	0.0090
			(0.0011)*	(0.0011)*	(0.0011)*	(0.0011)*
Job Group	-	-	-0.0135	-0.0135	-0.0135	-0.0128
			(0.0015)*	(0.0015)*	(0.0015)*	(0.0015)*
Manager	-	-	0.0703	0.0703	0.0697	0.0651
			(0.0062)*	(0.0062)*	(0.0062)*	(0.0062)*
Unemployment	-	-	-0.0651	-0.0654	-0.0645	-0.0415
			(0.0124)*	(0.0124)*	(0.0124)*	(0.0120)*
Benefits	-	-	-0.0000	-0.0000	-0.0000	-0.0000
			(0.0000)*	(0.0000)*	(0.0000)*	(0.0000)*
Constant	5.5772	6.4055	6.3153	6.3241	6.3102	26.7381
	(0.0236)*	(0.4078)*	(0.4684)*	(0.4681)*	(0.4684)*	(0.5519) *
Observations	133,898	124,385	109,244	109,244	109,244	109,284
Adj. R-Sq.	0.2880	0.8850	0.8908	0.8908	0.8908	0.8882
R-Sq. Within	-	0.0022	0.0201	0.0201	0.0205	0.0782

Note: * and ** show significance at 1% and 5% levels, respectively. Columns 1-5 results are for gross nominal wage as dependent variable. Column 6 gives result for gross real wage. The R-square for OLS is unadjusted. Standard errors are given in parenthesis.

Table C.3: Two-Way Fixed-Effects Results from the BHPS Panel

Variables	*OLS*	*FE*	*FE*	*FE*	*FE*	*FE*
Train Intensity	-0.0000	-0.0000	-0.0000	-	-0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)		(0.0000)	(0.0000)
Train Incidence	0.0214	0.0075	0.0067	-	-	0.0068
	(0.0008)*	(0.0014)*	(0.0014)*			(0.0013)*
Workplace Train	-	-	-	-	0.0044	-
					(0.0012)*	
Train Intensity2	-	-	-	-0.0000	-	-
				(0.0000)		
Train Incidence2	-	-	-	0.0041	-	-
				(0.0011)*		
Age	0.0679	0.1812	0.1634	0.1608	0.1643	0.4843
	(0.0059)*	(0.0375)*	(0.0371)*	(0.0371)*	(0.0372)*	(0.0379)*
Age Square	-0.0007	-0.0037	-0.0033	-0.0032	-0.0033	-.0066
	(0.0002)*	(0.0009)*	(0.0009)*	(0.0009)*	(0.0009)*	(0.0009)*
Age Cube	-0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	(0.0000)*	(0.0000)*	(0.0000)**	(0.0000)**	(0.0000)**	(0.0000)
Qualification	-0.2497	-0.1858	-0.2339	-0.2315	-0.2322	-0.2337
	(0.0173)*	(0.0831)**	(0.0805)*	(0.0806)*	(0.0805)*	(0.0792)*
Qualif. Square	0.0146	0.0406	0.0533	0.0523	0.0524	0.0530
	(0.0047)*	(0.0217)	(0.0209)**	(0.0209)**	(0.0209)**	(0.0205)*
Qualif. Cube	0.0000	-0.0026	-0.0034	-0.0034	-0.0034	-0.0034
	(0.0003)	(0.0015)	(0.0014)**	(0.0014)**	(0.0014)**	(0.0014)**
Firm Size	-	-	0.0311	0.0311	0.0312	0.0300
			(0.0062)*	(0.0062)*	(0.0062)*	(0.0060)*
F. Size Square	-	-	-0.0020	-0.0020	-0.0020	-0.0020
			(0.0006)*	(0.0006)*	(0.0006)*	(0.0005)*
Job Tenure	-	-	0.0016	0.0016	0.0016	0.0011
			(0.0005)*	(0.0005)*	(0.0005)*	(0.0005)**
Permanent	-	-	0.0817	0.0817	0.0821	0.0866
			(0.0156)*	(0.0157)*	(0.0156)*	(0.0150)*
Second Job	-	-	-0.0389	-0.0390	-0.0388	-0.0333
			(0.0088)*	(0.0088)*	(0.0088)*	(0.0086)*
Satisfaction	-	-	0.0108	0.0107	0.0108	0.0101
			(0.0015)*	(0.0015)*	(0.0015)*	(0.0015)*
Unemployment	-	-	-0.0558	-0.0558	-0.0559	-0.0588
			(0.0111)*	(0.0111)*	(0.0111)*	(0.0106)*
Other Income	-	-	-0.0000	-0.0000	-0.0000	-0.0000
			(0.0000)*	(0.0000)*	(0.0000)*	(0.0000)
Constant	6.3310	4.8730	4.9046	4.9470	4.9003	-1.5922
	(0.0754)*	(0.5178)*	(0.5151)*	(0.5150)*	(0.5154)*	(0.5312)*
Observations	57,531	54,529	53,894	53,894	53,894	53,894
Adj. R-Sq.	0.2098	0.8983	0.9001	0.9000	0.9000	0.9033
R-Sq. Within	-	0.0032	0.0148	0.0144	0.0144	0.0483

Note: * and ** show significance at 1% and 5% levels, respectively. Columns 1-5 results are for gross nominal wage as dependent variable. Column 6 gives result for gross real wage. The R-square for OLS is unadjusted. Standard errors are given in parenthesis.

Table C.4: Training Effect Estimates with Abraham and Sun (2018) from UKHLS Subpanel

r' relative to first training	FE $Estimtes$	IWD $Estimtes$	Estimates for ATT for the five cohorts Coh_1	Coh_2	Coh_3	Coh_4	Coh_5
-4	-0.018	-0.003	-	-	-	-	-0.008
	(0.036)	(0.013)					(0.038)
-3	0.012	0.008	-	-	-	0.048	-0.004
	(0.059)	(0.007)				(0.029)**	(0.036)
-2	-0.007	0.003	-	-	-0.005	0.039	-0.046
	(0.083)	(0.010)			(0.027)	(0.031)	(0.037)
-1	0.012	0.016	-	0.039	-0.000	0.003	0.068
	(0.109)	(0.016)		(0.030)	(0.029)	(0.032)	(0.049)
0	0.019	0.0260	0.020	0.045	-0.024	0.069	-0.001
	(0.134)	(0.025)	(0.028)	(0.032)	(0.034)	(0.046)	(0.041)
1	0.029	0.033	0.034	0.021	0.043	0.039	-
	(0.160)	(0.026)	(0.029)	(0.032)	(0.045)	(0.037)	
2	0.032	0.022	0.001	0.076	-0.007	-	-
	(0.186)	(0.029)	(0.030)	(0.046)**	(0.036)		
3	0.053	0.069	0.088	0.039	-	-	-
	(0.212)	(0.036)*	(0.044)*	(0.038)			
4	0.051	0.025	0.025	-	-	-	-
	(0.237)	(0.034)	(0.034)				
5	0.060	-	-	-	-	-	-
	(0.263)						
Observations	17,087		14,646	14,646	14,646	14,646	14,646
Adjusted R-Sq.	0.8669		0.8712	0.8712	0.8712	0.8712	0.8712
R-Sq. Within	0.0924		0.0793	0.0793	0.0793	0.0793	0.0793

Note: Only those individuals are part of this subpanel who have data for all waves and are trained at least once in the sample period. Significance at 5% and 10% levels is indicated by * and **, respectively.

C.3 Figures

1. $ATT(e,t)$ from Callaway and Sant'Anna (2018) for the UKHLS

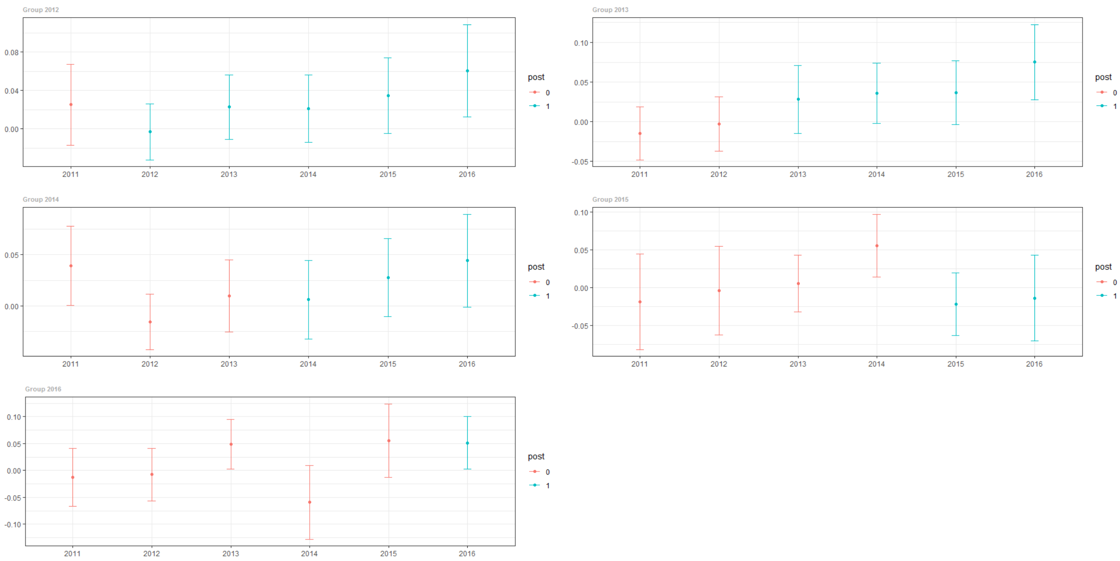

Figure C.1: Un-Conditional DID Estimations Based on UKHLS

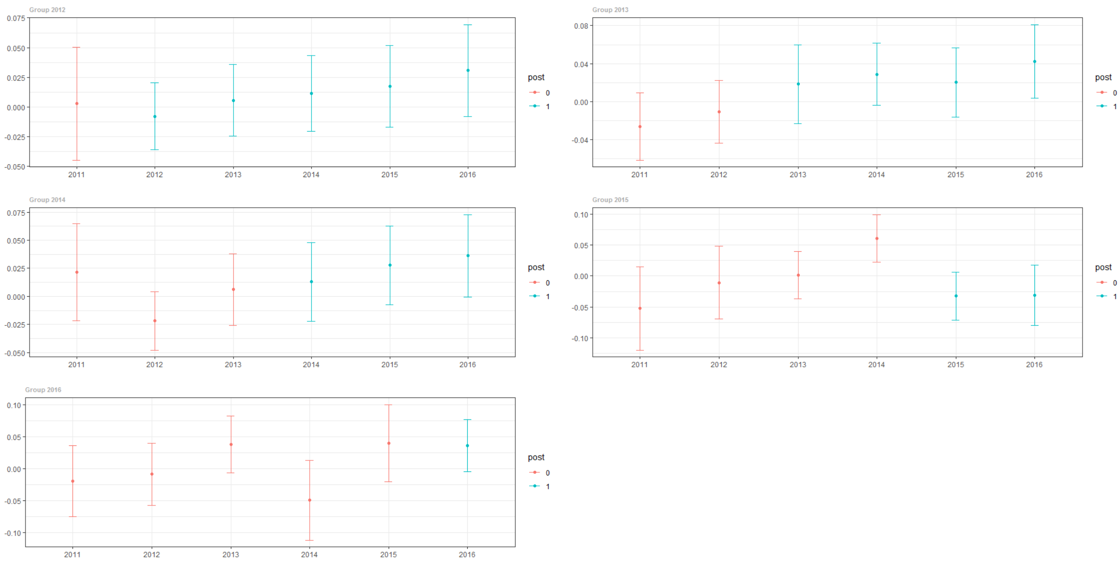

Figure C.2: Conditional DID Estimations based on UKHLS

2. $ATT(L, F)$ from Imai et al. (2018)

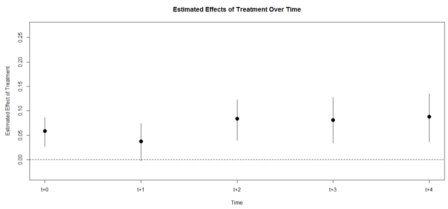

Figure C.3: Column 1 of Table 4.3

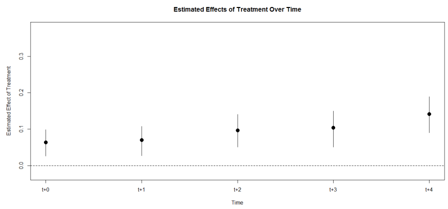

Figure C.4: Column 2 of Table 4.3

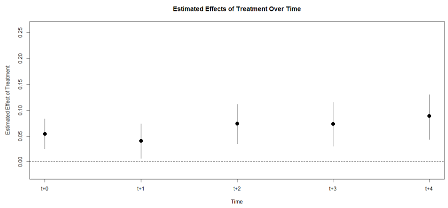

Figure C.5: Column 4 of Table 4.3

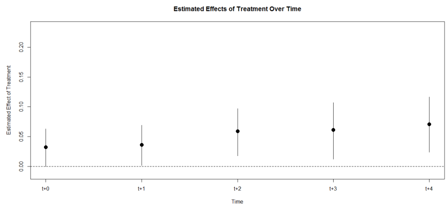

Figure C.6: Column 5 of Table 4.3

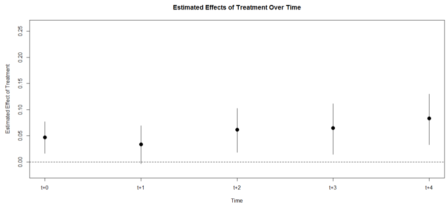

Figure C.7: Column 6 of Table 4.3

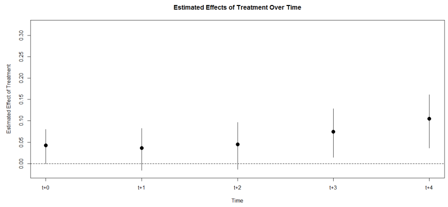

Figure C.8: Column 7 of Table 4.3

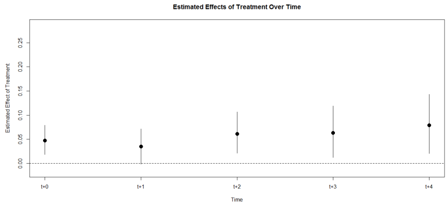

Figure C.9: Column 8 of Table 4.3

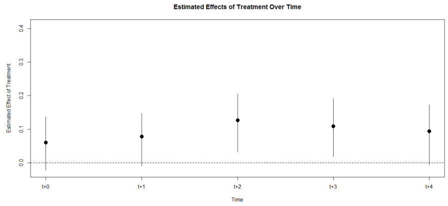

Figure C.10: Column 9 of Table 4.3

Bibliography

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *The Review of Economic Studies*, 72(1):1–19.

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

Abraham, S. and Sun, L. (2018). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Papers 1804.05785, arXiv.org.

Acemoglu, D. and Pischke, J.-S. (1996). Why do firms train? theory and evidence. Working Paper 5605, National Bureau of Economic Research.

Acemoglu, D. and Pischke, J.-S. (1998). Why do firms train? theory and evidence. *The Quarterly Journal of Economics*, 113(1):79–119.

Acemoglu, D. and Pischke, J.-S. (1999). The structure of wages and investment in general training. *Journal of political economy*, 107(3):539–572.

Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, 83(6):2411–2451.

Aghion, P., Howitt, P., and Violante, G. (2002). General purpose technology and wage inequality. *Journal of Economic Growth*, 7(4):315–45.

Akerlof, G. A. and Yellen, J. L. (1990). The fair wage-effort hypothesis and unemployment. *The Quarterly Journal of Economics*, 105(2):255–283.

Albert, C., García-Serrano, C., and Hernanz, V. (2010). On-the-job training in europe: Determinants and wage returns. *International Labour Review*, 149(3):315–341.

Aloi, M. and Tournemaine, F. (2013). Inequality, growth, and environmental quality trade-offs in a model with human capital accumulation. *Canadian Journal of Economics/Revue canadienne d'économique*, 46(3):1123–1155.

Altonji, J. G. and Spletzer, J. R. (1991). Worker characteristics, job characteristics, and the receipt of on-the-job training. *ILR Review*, 45(1):58–79.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Arulampalam, W., Booth, A. L., and Bryan, M. L. (2004). Training in europe. *Journal of the European Economic Association*, 2(2-3):346–360.

Arulampalam, W. N. and Booth, A. L. (2001). Learning and earning: Do multiple training events pay? a decade of evidence from a cohort of young british men. *Economica*, 68(271):379–400.

Athey, S. and Imbens, G. W. (2018). Design-based analysis in difference-in-differences settings with staggered adoption. Technical report, National Bureau of Economic Research.

Autor, D. H. (2014). Skills, education, and the rise of earnings inequality among the “other 99 percent”. *Science*, 344(6186):843–851.

Barron, J. M., Berger, M. C., and Black, D. A. (1999). Do workers pay for on-the-job training? *The Journal of Human Resources*, 34(2):235–252.

Barron, J. M., Black, D. A., and Loewenstein, M. A. (1989). Job matching and on-the-job training. *Journal of Labor Economics*, 7(1):1–19.

Bartel, A. P. and Sicherman, N. (1998). Technological change and the skill acquisition of young workers. *Journal of Labor Economics*, 16(4):718–755.

Barth, E., Bryson, A., Davis, J. C., and Freeman, R. (2016). It’s where you work: Increases in the dispersion of earnings across establishments and individuals in the united states. *Journal of Labor Economics*, 34(S2):S67–S97.

Bartlett, K. R. (2001). The relationship between training and organizational commitment: A study in the health care field. *Human resource development quarterly*, 12(4):335–352.

Bassanini, A. and Ok, W. (2004). How do firms’ and individuals’ incentive to invest in human capital vary across groups? Technical report, OECD Publishing.

Becker, G. (1975). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education, Second Edition*. National Bureau of Economic Research, Inc.

Becker, G. (1981). *Treatise on the Family*. Harvard University Press, Cambridge.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of political economy*, 70(5, Part 2):9–49.

Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *Journal of Political Economy*, 75(4, Part 1):352–365.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, 119(1):249–275.

Black, D. A. and Loewenstein, M. A. (1997). Dismissals and match-specific rents. *Labour Economics*, 4(4):325–340.

Blundell, R., Dearden, L., and Meghir, C. (1996). *The determinants and effects of work-related training in Britain*. Number R50. IFS Reports, Institute for Fiscal Studies.

Blundell, R., Dearden, L., and Meghir, C. (1999). Work-related training and earnings. In *ZEW Conference on Programme Evaluation, June*.

Booth, A. L. (1991). Job-related formal training: Who receives it and what is it worth? *Oxford bulletin of economics and statistics*, 53(3):281–294.

Booth, A. L. and Bryan, M. L. (2002). Who pays for general training? new evidence for british men and women. IZA Discussion Papers 486, Institute for the Study of Labor (IZA).

Booth, A. L. et al. (1993). Private sector training and graduate earnings. *The Review of Economics and Statistics*, 75(1):164–170.

Booth, A. L. and Zoega, G. (2004). Is wage compression a necessary condition for firm-financed general training? *Oxford Economic Papers*, 56(1):88–97.

Borusyak, K. and Jaravel, X. (2017). Revisiting event study designs. *Available at SSRN 2826228*.

Braveman, P. A., Cubbin, C., Egerter, S., Williams, D. R., and Pamuk, E. (2010). Socioeconomic disparities in health in the united states: What the patterns tell us. *American journal of public health*, 100(S1):S186–S196.

Brown, C. and Medoff, J. (1989). The employer size-wage effect. *Journal of Political Economy*, 97(5):1027–1059.

Brunello, G. (2001). On the complementarity between education and training in Europe. IZA Discussion Papers 309, Institute for the Study of Labor (IZA).

Callaway, B. and Sant'Anna, P. H. C. (2018). Difference-in-differences with multiple time periods and an application on the minimum wage and employment. DETU Working Papers 1804, Department of Economics, Temple University.

Card, D., Heining, J., and Kline, P. (2013). Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly Journal of Economics*, 128(3):967–1015.

Cohn, A., Fehr, E., and Goette, L. (2014). Fair wages and effort provision: Combining evidence from a choice experiment and a field experiment. *Management Science*, 61(8):1777–1794.

Colombo, E. and Stanca, L. (2008). The impact of training on productivity: Evidence from a large panel of firms. Working Paper 134, University of Milano-Bicocca, Department of Economics.

Conti, G. (2005). Training, productivity and wages in italy. *Labour economics*, 12(4):557–576.

Costen, W. M. and Salazar, J. (2011). The impact of training and development on employee job satisfaction, loyalty, and intent to stay in the lodging industry. *Journal of Human Resources in Hospitality & Tourism*, 10(3):273–284.

Davis, S. J. and Haltiwanger, J. (1991). Wage dispersion between and within u.s. manufacturing plants, 1963-1986. Working Paper 3722, National Bureau of Economic Research.

de Chaisemartin, C. and D'Haultfœuille, X. (2018). Two-way fixed effects estimators with heterogeneous treatment effects. *arXiv preprint arXiv:1803.08807*.

Dearden, L., Reed, H., and Van Reenen, J. (2006). The impact of training on productivity and wages: Evidence from british panel data. *Oxford bulletin of economics and statistics*, 68(4):397–421.

Dietz, D. and Zwick, T. (2016). The retention effect of training - portability, visibility, and credibility. Discussion Paper 16-011, ZEW - Centre for European Economic Research.

Dobkin, C., Finkelstein, A., Kluender, R., and Notowidigdo, M. J. (2018). The economic consequences of hospital admissions. *American Economic Review*, 108(2):308–52.

Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2009). Homo reciprocans: Survey evidence on behavioural outcomes. *The Economic Journal*, 119(536):592–612.

Englmaier, F., Kolaska, T., and Leider, S. (2015). Reciprocity in organisations - evidence from the wers. Working Paper 5168, CESifo.

Felstead, A., Gallie, D., Green, F., and Inanc, H. (2013). Skills at work in britain - first findings from the skills and employment survey 2012.

Felstead, A., Gallie, D., Green, F., and Zhou, Y. (2007). *Skills at work in Britain, 1986 to 2006*. ESRC Centre on Skills, Knowledge and Organisational Performance.

Fouarge, D., Schils, T., and de Grip, A. (2013). Why do low-educated workers invest less in further training? *Applied Economics*, 45(18):2587–2601.

Frazis, H. and Loewenstein, M. A. (2005). Reexamining the returns to training functional form, magnitude, and interpretation. *Journal of Human Resources*, 40(2):453–476.

Frazis, H. J. and Spletzer, J. R. (2005). Worker training: What we've learned from the nlsy79. *Monthly Labor Review*, 128:48–58.

Friedrich, B. (2016). Internal labor markets and the competition for managerial talent. Technical report, Working Paper.

Gibbons, C. E., Suárez Serrato, J. C., and Urbancic, M. B. (2014). Broken or fixed effects? Working Paper 20342, National Bureau of Economic Research.

Goodman-Bacon, A. (2018). Difference-in-differences with variation in treatment timing. Working Paper 25018, National Bureau of Economic Research.

Görlitz, K. (2011). Continuous training and wages: An empirical analysis using a comparison-group approach. *Economics of Education Review*, 30(4):691 – 701.

Görlitz, K. and Tamm, M. (2016). The returns to voucher-financed training on wages, employment and job tasks. *Economics of Education Review*, 52:51 – 62.

Green, F. (1993). The determinants of training of male and female employees in britain. *Oxford bulletin of economics and statistics*, 55(1):103–122.

Haelermans, C. and Borghans, L. (2012). Wage effects of on-the-job training: A meta-analysis. *British Journal of Industrial Relations*, 50(3):502–528.

Havranek, T., Horvath, R., Irsova, Z., and Rusnak, M. (2015). Cross-country heterogeneity in intertemporal substitution. *Journal of International Economics*, 96(1):100 – 118.

He, H. (2012). What drives the skill premium: Technological change or demographic variation? *European Economic Review*, 56(8):1546–1572.

Heckman, J. J. (1976). A life-cycle model of earnings, learning, and consumption. *Journal of Political Economy*, 84(4):S9–S44.

Heckman, J. J., Lochner, L., and Taber, C. (1998). Explaining rising wage inequality: Explorations with a dynamic general equilibrium model of labor earnings with heterogeneous agents. *Review of Economic Dynamics*, 1(1):1 – 58.

Henseke, G., Felstead, A., Gallie, D., and Green, F. (2018). Skills trends at work in britain - first findings from the skills and employment survey 2017.

Hidalgo, D., Oosterbeek, H., and Webbink, D. (2014). The impact of training vouchers on low-skilled workers. *Labour Economics*, 31:117–128.

Imai, K. and Kim, I. S. (2019a). On the use of two-way fixed effects regression models for causal inference with panel data. *Unpublished Manuscript*.

Imai, K. and Kim, I. S. (2019b). When should we use unit fixed effects regression models for causal inference with longitudinal data? *American Journal of Political Science*.

Imai, K., Kim, I. S., and Wang, E. (2018). Matching methods for causal inference with time-series cross-sectional data.

Jones, D. C., Kalmi, P., and Kauhanen, A. (2011). The effects of general and firm-specific training on wages and performance: Evidence from banking. *Oxford economic papers*, 64(1):151–175.

José Chambel, M. and Sobral, F. (2011). Training is an investment with return in temporary workers: A social exchange perspective. *Career Development International*, 16(2):161–177.

Kanervisto, M., Vasankari, T., Laitinen, T., Heliövaara, M., Jousilahti, P., and Saarelainen, S. (2011). Low socioeconomic status is associated with chronic obstructive airway diseases. *Respiratory Medicine*, 105(8):1140 – 1146.

Katz, E. and Ziderman, A. (1990). Investment in general training: The role of information and labour mobility. *The Economic Journal*, 100(403):1147–1158.

Kessler, A. S. and Lülfesmann, C. (2006). The theory of human capital revisited: On the interaction of general and specific investments. *The Economic Journal*, 116(514):903–923.

King, R. and Rebelo, S. (1999). Resuscitating real business cycles. In Taylor, J. B. and Woodford, M., editors, *Handbook of Macroeconomics*, volume 1, Part B, chapter 14, pages 927–1007. Elsevier, 1 edition.

Konings, J. and Vanormelingen, S. (2015). The impact of training on productivity and wages: Firm-level evidence. *Review of Economics and Statistics*, 97(2):485–497.

Kube, S., Maréchal, M. A., and Puppe, C. (2012). The currency of reciprocity: Gift exchange in the workplace. *American Economic Review*, 102(4):1644–62.

Kuckulenz, A. and Zwick, T. (2003). The impact of training on earnings: Differences between participant groups and training forms. Technical report, ZEW Discussion Papers.

Lazear, E. P. (2009). Firm-specific human capital: A skill-weights approach. *Journal of political economy*, 117(5):914–940.

Leamer, E. E. (1999). Effort, wages, and the international division of labor. *Journal of Political Economy*, 107(6):1127–1162.

Lengermann, P. A. (2000). How long do the benefits of training last? evidence of long term effects across current and previous employers. In *Research in Labor Economics*, pages 439–461. Emerald Group Publishing Limited.

Leuven, E. and Oosterbeek, H. (2000). The demand and supply of work-related training: Evidence from four countries. In *Research in labor economics*, pages 303–330. Emerald Group Publishing Limited.

Leuven, E. and Oosterbeek, H. (2004). Evaluating the effect of tax deductions on training. *Journal of Labor Economics*, 22(2):461–488.

Leuven, E. and Oosterbeek, H. (2008). An alternative approach to estimate the wage returns to private-sector training. *Journal of applied econometrics*, 23(4):423–434.

Leuven, E., Oosterbeek, H., Sloof, R., and Van Klaveren, C. (2005). Worker reciprocity and employer investment in training. *Economica*, 72(285):137–149.

Lillard, L. A. and Tan, H. W. (1986). *Introduction to "Private Sector Training: Who gets it and what are its Effects?"*. ERIC.

Loewenstein, M. A. and Spletzer, J. (1999). General and specific training: Evidence and implications. *Journal of Human Resources*, 34(4):710–733.

Loewenstein, M. A. and Spletzer, J. R. (1998). Dividing the costs and returns to general training. *Journal of Labor Economics*, 16(1):142–171.

Lubetkin, E. I. and Jia, H. (2017). Burden of disease associated with lower levels of income among us adults aged 65 and older. 7(1):e013720.

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, 22(1):3 – 42.

Lynch, L. M. (1991). The role of off-the-job vs. on-the-job training for the mobility of women workers. *The American Economic Review*, 81(2):151–156.

Lynch, L. M. (1992). Private-sector training and the earnings of young workers. *The American Economic Review*, 82(1):299–312.

Maximiano, S. and Oosterbeek, H. (2007). On the determinants of workers' and firms' willingness to train. In *Conference paper, European Association of Labour Economists (EALE), Oslo.*

Mincer, J. (1974a). *Schooling, Experience, and Earnings.* National Bureau of Economic Research, Inc.

Mincer, J. A. (1974b). Schooling and earnings. In *Schooling, experience, and earnings*, pages 41–63. NBER.

Munandar, H. (2008). Heterogeneous agents, human capital formation and international income inequality. Tinbergen Institute Discussion Papers 08-015/2, Tinbergen Institute.

NCHS (2012). Health in united states, 2011: With special feature on socioeconomic status and health.

OECD (1991). *Employment Outlook. Enterprise-Related Training.* OECD Publishing, Paris.

OECD (1993). *Employment Ooutlook 1993.* OECD Publishing, Paris.

OECD (1999). *Employment Outlook. Training of Adults Workers in OECD Countries: Measurement and Analysis.* OECD Publishing, Paris.

OECD (2003). *Employment Outlook. Upgrading Workers' Skills and Competencies.* OECD Publishing, Paris.

OECD (2004). *Co-financing lifelong learning: Towards a systemic approach.* OECD Publishing, Paris.

OECD (2005). *Promoting Adult Learning.* OECD Publishing, Paris.

OECD (2010). *Tackling Inequalities in Brazil, China, India and South Africa: The Role of Labour Market and Social Policies.* OECD Publishing.

Oi, W. Y. and Idson, T. L. (1999). Chapter 33 firm size and wages. volume 3 of *Handbook of Labor Economics*, pages 2165 – 2214. Elsevier.

Parent, D. (1999). Wages and mobility: The impact of employer-provided training. *Journal of labor economics*, 17(2):298–317.

Parent, D. (2003). Employer-supported training in canada and its impact on mobility and wages. *Empirical Economics*, 28(3):431–459.

Peach, E. K. and Stanley, T. D. (2009). Efficiency wages, productivity and simultaneity: A meta-regression analysis. *Journal of Labor Research*, 30(3):262–268.

Pew Research Center (2016). The state of american jobs.

Pischke, J.-S. (2001). Continuous training in germany. *Journal of population economics*, 14(3):523–548.

Pollack, C. E., Cubbin, C., Sania, A., Hayward, M., Vallone, D., Flaherty, B., and Braveman, P. A. (2013). Do wealth disparities contribute to health disparities within racial/ethnic groups? *J Epidemiol Community Health*, 67(5):439–445.

Rebelo, S. (1991). Long-Run Policy Analysis and Long-Run Growth. *Journal of Political Economy*, 99(3):500–521.

Rebelo, S. et al. (1998). *The role of knowledge and capital in economic growth*. Number 149. UNU World Institute for Development Economics Research.

Rotemberg, J. J. (2006). Altruism, reciprocity and cooperation in the workplace. *Handbook of the economics of giving, altruism and reciprocity*, 2:1371–1407.

Rowden, R. W. and Conine Jr, C. T. (2005). The impact of workplace learning on job satisfaction in small us commercial banks. *Journal of workplace Learning*, 17(4):215–230.

Sauermann, J. (2015). Worker reciprocity and the returns to training: Evidence from a field experiment. IZA Discussion Papers 9179, Institute for the Study of Labor (IZA).

Schmidt, S. W. (2007). The relationship between satisfaction with workplace training and overall job satisfaction. *Human Resource Development Quarterly*, 18(4):481–498.

Schøne, P. (2004). Why is the return to training so high? *Labour*, 18(3):363–378.

Schwerdt, G., Messer, D., Woessmann, L., and Wolter, S. C. (2012). The impact of an adult education voucher program: Evidence from a randomized field experiment. *Journal of Public Economics*, 96(7):569 – 583.

Sousounis, P. (2009). The impact of work-related training on employee earnings: Evidence from great britain. Mpra paper, University Library of Munich, Germany.

Stevenson, B. and Wolfers, J. (2006). Bargaining in the shadow of the law: Divorce laws and family distress. *The Quarterly Journal of Economics*, 121(1):267–288.

Verhaest, D. and Omey, E. (2013). The relationship between formal education and skill acquisition in young workers' first jobs. *The Manchester School*, 81(4):638–659.

Veum, J. R. (1995a). Sources of training and their impact on wages. *ILR Review*, 48(4):812–826.

Veum, J. R. (1995b). Training wages and the human capital model. [Unpublished] 1995. Presented at the Annual Meeting of the Population Association of America San Francisco California April 6-8 1995.

Veum, J. R. (1997). Training and job mobility among young workers in the united states. *Journal of population economics*, 10(2):219–233.

Waldman, M. (1984). Job assignments, signalling, and efficiency. *The RAND Journal of Economics*, 15(2):255–267.

Waldman, M. (1990). Up-or-out contracts: A signaling perspective. *Journal of Labor Economics*, 8(2):230–250.

Winterbotham, M., Vivian, D., Kik, G., Huntley, J., Hewitt, M. T., Downing, C., Thomson, D., Morrice, N., and Stroud, S. (2018). Employer skills survey 2017.

Woolf, S., Aron, L., Dubay, L., Zimmerman, E., Simon, S., and Lux, K. (2015). *How are Income and Wealth Linked to Health and Longevity?* Income and health initiative. Virginia Commonwealth University, Center on Society and Health.