Hitotsubashi Journal of Economics 51 (2010), pp.79-97.

INITIAL WAGE, HUMAN CAPITAL AND POST WAGE DIFFERENTIALS

Li Gan

Department of Economics, Texas A&M University
College Station, TX 77843-4228, USA
gan@econmail.tamu.edu.

Jaeun Shin*

KDI School of Public Policy and Management
Seoul 130-868, Korea
jshin@kdischool.ac.kr.

Qi Li

Department of Economics, Texas A&M University
College Station, TX 77843-4228, USA
qi@econmail.tamu.edu.

Received April 2009; Accepted June 2010

Abstract

Insufficiency in information with which firms judge the productivity of a worker for the first time in the market creates more randomness in initial wages than in later wages. This paper examines whether the initial randomness in wages may have a persistent effect on post wages. We set up a human capital accumulation in which an individual may respond to the positive error in initial wage by adjusting hours worked thereafter in her career, and consequently may receive higher future wages than those who draw a negative error in initial wages but otherwise are equivalent. The model predicts that the initial wage, in particular, its random component, is a persistently important factor having positive effect on future wages. Using data from the National Longitudinal Survey of Youth 79, we find empirical evidence that this effect is indeed positive and persists even after 20 years since the initial entry to labor market. The decomposition of initial wages by both parametric and nonparametric IV methods further shows that this effect is derived by the random component, not the observable component, of the initial wage. It implies that the observed cross-sectional wage variation within group can be accounted for the initial randomness in wages.

Key words: Human capital accumulation; Learning; Initial wage; Wage differentials
JEL Classification: J24; J31; C14

---

* Corresponding author.

I. *Introduction*

One of the most robust findings in labor economics is a positive return on time spent in the labor market investigated in the standard log-wage equations. Among various attempts to explain this relationship, the primary model is the general human capital theory, in which the stock of human capital rises with experience in the labor market. There are two main streams to describe how people accumulate their human capital through working. First, Ben-Porath (1967) and many followers (Mincer, 1974; Becker, 1975) suggest that people invest in general human capital at the expense of time or direct pecuniary cost, resulting initially in lower wages and subsequently in higher wages in later periods of working life. This so-called 'On the Job Training' (OJT) theory of post-schooling human capital investment and wage growth predicts that at the individual level, there will be a negative relationship between the initial wage level and wage growth. At the market level, it also predicts that the present values of the investor's lifetime wage (with lower initial wage and faster growth) and of the otherwise equivalent non-investor's lifetime wage (with higher initial wage but slower growth thereafter) equals one. These two predictions are tested to be valid in the empirical study of Neumark and Taubman (1995).

The other hypothesis regarding the general human capital accumulation process, namely 'Learning by Doing' (LBD) theory, holds that people enhance their productivity by learning skills through work without giving up working hours or wages. In the two-period model (Cossa, Heckman, and Lochner, 1999), workers choose optimal consumption and leisure (in hours) to maximize their present value lifetime utility. This gives different predictions from the OJT and the LBD human capital evolution processes on the relationships among initial wage level, hours worked, and wage growth over time. In the OJT, people reduce hours worked for post-schooling training and receive lower wages initially. Over time, their wages grow faster. Therefore, their wages and hours worked move in opposite directions over time. People who invest in human capital earn lower initial wages and experience higher growth in their future wages. The gap in the initial wages would be compensated by difference in wage growth rates. As age approaches the overtaking point, two wage profiles (one flatter than the other) may converge to each other (Mincer, 1974).

By contrast, in the LBD process, there is no explicit cost for human capital accumulation. Given wage offers, people choose how many hours to work and by working, they enhance their productivity. As wages reflect increases in their stock of human capital, wages increase over time as well, and this increase will be larger for workers who spend more hours at work. Since people adjust their labor supply (measured in hours) responding to wages, the patterns of wage and hours worked would move in parallel with each other; the higher the wages, the more hours they work. Moreover, assuming no costly training and flexible choice in hours worked, a difference in initial wages may cause different choices of hours worked, and so, different amount of accumulated human capital. Eventually, the small gap in initial wages is transmitted to a huge wage differential in the later working years.

Empirical evidence such as Neumark and Taubman (1995) has limitations for verifying that their findings indeed prove the OJT hypothesis because they have not considered the endogenous training participation decisions of workers explicitly. When they find a negative relationship between wage level and wage growth, it is simply argued to be evidence for the

OJT. The separate identification and reconciliation between the OJT and LBD models are little examined.

Aiming to provide evidence for the LBD model, our study considers the initial wage effect on post wage differentials. Variation in initial wages is derived from two factors; a worker-specific permanent factor and random transitory components. The fact that people with higher initial wages earn more in later periods than their counterparts with lower initial wages may be attributed to the possibility that people with higher initial wages are smarter (endowed with more human capital). Then, differences in post initial wages are consequences in their difference in innate ability before working. If the growth rate of human capital accumulation is independent of (dependent on) individual endowed or prior-to-working ability, the gap in initial wages will remain as the same extent (increase in the magnitude) in future wages without any slope effect (with positive slope effect). This implies that the permanent components of initial wages are mostly responsible for future wage differentials.

By decomposing initial level wages into their permanent and transitory components (Baker, 1997), the relative accountability of these two components for post wage differentials can be investigated. It is expected that the determination of initial wage is quite different from that of later wages because when a firm decides to hire someone who enters the market for the first time, the firm usually does not have enough information to accurately judge the productivity of the individual. Therefore, there is more randomness in assigning an initial wage compared with later wages.[1]

In sum, there are two possible ways in which initial wages may influence future wage: (1) higher initial wages increase hours worked (labor supply effect); and (2) higher initial wages cause faster wage growth (slope effect). This paper mainly focuses on the slope effect of initial wages on future wages and pays special attention to the role of transitory components of initial wage in determining future wages. It is found that the initial wage is indeed an important factor in determining future wages both in the theoretical framework and in the empirical analysis.

We first model the human capital accumulation process in the LBD framework , in which an individual worker is endowed with innate ability. She adjusts labor supply decisions on hours worked in response to her initial wage. With the LBD, she accumulates her productivity proportional to hours worked. If the level of the starting wage is high, she works for marginally more hours in the next period. This leads her to obtain more human capital during working. Her future wage is likely to be high, directly reflecting her enhanced productivity. Thus, the positive initial randomness in wages may lead to a persistent future high wage by motivating the individual to work harder in order to keep a well-paid job in her career.

We further seek for empirical evidence on whether any of these effects is attributed to a permanent factor or to random shock in the initial wages. Using a simple regression model of the logarithm of wages applied to data from the National Longitudinal Survey of Youth 79, we estimate the effects of initial wage and its random component. To capture the effect of initial randomness in wage determination uncompounded from the role of innate ability, we

[1] The composite of factors that influence wages (or wage offers) that are beyond the worker's control, for example, firm-specific idiosyncratic preferences for workers traits is often defined as "luck" (Betrand, Mullainathan, 2001; Garvey, Milbourn 2006). An individual is *lucky* if she obtains a relatively high wage to others who have similar productivity. A few studies have shown that the payment to employees at executive level is sensitive to exogenous forces, that is, luck, as much as it is to firm-specific performance.

approximate the productivity of an individual using age, years of education, and AFQT scores. In addition, we perform the parametric IV estimation using parents' and the closest siblings' level of education and further conduct the nonparametric IV estimation developed by Newey (1990). The primary purpose of these instrumental approaches is to detangle the random component from the innate ability of an individual. Our estimation results consistently provide strong evidence that the initial wage has a positive effect on future wages and that it persists over time. This effect is found to be attributable to the *random* component of the initial wages, not the component based on observed individual attributes. This supports the positive relationship between wage level and wage growth predicted by the LBD model.

II. *Theoretical Framework*

Let a representative worker endowed with innate ability α_0, which is not fully observed by firms. We assume that once a worker is actually employed, then her ability becomes public information both to her employer and to the firm that she works for. A worker accumulates her human capital through her working experience in the LBD process. The wage paid by the firm to this representative worker is, on the average, equal to the observed output of a worker. The observed output can be decomposed into two parts: worker's productivity, and a random (or luck) component due to the firm's inability to observe the individual's true ability before hiring her.

Then, wages at time t are determined by the firm's expectation about the marginal product of a worker, α_t^e, plus a random noise ε_t:

$$w_t = \alpha_t^e + \varepsilon_t. \tag{1}$$

We let expected the marginal product of a worker at t be the marginal product of the worker at the end of *t-1*:

$$\alpha_t^e = \alpha_{t-1}. \tag{2}$$

Further we let a worker adjust her working hours, h_t, responding to her wage as follows:

$$h_t = h_0 + \beta(w_t - \alpha_{t-1}) = h_0 + \beta\varepsilon_t. \tag{3}$$

Equation (3) says that working hours by which we approximate total effort input at work is correlated with the difference between the wage and the marginal productivity, that is, the random component of wages, ε_t. The response of hours of work to ε_t is assumed to be constant, denoted as β. The exact sign and magnitude of β would be determined as the net of income effect and substitution effect of the ε_t. In this model, we postulate the positive sign for β because the opportunity cost of leisure gets increased due to the LBD benefit of working: additional leisure hour requires not only the loss of wage income but also the loss in human capital accumulation. The worker's incentive to accumulate human capital by working, and thus to earn higher post wages may lead her to take advantage of this favorable chance of the positive random shock in wages. Then, she is likely to invest more time on working in response to higher wage than her actual human capital (substitution effect dominates income effect).

In the line of the Learning-by-doing human capital theory (Cossa, Heckman, and Lochner, 1999; Heckman, Lochner, and Cossa, 2002), we model the accumulation process of a worker's ability to depend on working hours as follows:

$$\alpha_t=\rho\alpha_{t-1}+\eta h_{t-1}, \tag{4}$$

where ρ measures the rate of human capital depreciation. If $\rho=1$, then Equation (4) implies that the human capital stock is never depreciated. Then through a lifetime, the productivity of a worker keeps increasing. The productivity of a worker is assumed to enhanced by the last period working at a constant learning rate of η. If η is zero, it means that no human capital accumulation is achieved from work ($\eta\geq 0$).

Simple deduction from Equations (3) and (4) leads us to:

$$\alpha_t=\rho^t\alpha_0+\frac{1-\rho^t}{1-\rho}\eta h_0+\eta\beta\rho^{t-1}\varepsilon_0+\eta\beta\sum_{j=1}^{t-1}\rho^{t-1-j}\varepsilon_j. \tag{5}$$

The first term on the right-hand-side (RHS) of (5) is the initial human capital of the worker. It is constructed from time-invariant characteristics such as years of schooling, sex, and race. The second term at the RHS of (5) is a time-varying component of the obtained productivity from experience in the workplace. It includes working experience and tenure with a specific employer. The third term, ε_0, represent a random component of the initial wage or *initial luck* , which is our main concern.

To predict the sign and trend of the initial wage effect, we consider four cases where different distributional assumptions on the random component of initial wages and variation from the benchmark process of human capital accumulation are considered. In particular,

Case 1. $\rho=1$, and $\{\varepsilon_t\}$ is *i.i.d.* with mean zero and finite variance σ^2 for t=1,...,T. We allow that ε_0 can have a much larger variation than σ^2. The human capital accumulation process follows Equation (4). Random shock in initial wage reflects the luck of an individual. With $\rho=1$, Equation (5) becomes

$$\alpha_t=\alpha_0+t\eta h_0+\eta\beta\varepsilon_0+\eta\beta\sum_{j=1}^{t-1}\varepsilon_j. \tag{6}$$

We predict that the initial random shock in wages has a permanent and positive impact on future human capital for all future time periods to the extent of $\eta\beta$.

In this case, we can interpret the error component in the initial wage as the quality of matching between a worker and an employer (or a workplace environment). In general, each worker only has limited access to information about job openings. A worker could afford to search and gather employment opportunity information only when she manages to be employed during the searching process or support herself while unemployed. So her matching with a job would occur at a time when she can not afford the searching process any more and she would choose a job from the offers available to her then. At any moment, it is not practically possible that a worker searches the best match using the entire information about the job market.

In the other case where a worker does not try to change her employer, her workplace environment with the same employer can still change randomly in the sense that many economy-wide factors that a worker is unable to control may affect her working environment relevant to her productivity. In either case, it is reasonable to assume that the matching quality is random (has a random component) over time.

Case 2. $0<\rho<1$, i.e. α_t follows a stationary *AR (1)* process given as:

$$\alpha_t=\rho\alpha_{t-1}+\eta h_0+\eta\beta\varepsilon_{t-1} \text{ where } 0<\rho<1. \quad (7)$$

Equation (7) allows the possibility of human capital depreciation over time. It is reasonable that as they get older, workers may forget part of their knowledge accumulated in the past. If they switch jobs and work in different tasks, part of the specific human capital obtained in previous jobs may be lost in their productivity in the next period. In either case, workers' human capital may depreciate for a given time period. Direct deduction from Equation (7) leads to:

$$\alpha_t=\rho^t\alpha_0+\frac{1-\rho^t}{1-\rho}\eta h_0+\eta\beta\rho^{t-1}\varepsilon_0+\eta\beta\sum_{j=1}^{t-1}\rho^{t-1-j}\varepsilon_j. \quad (8)$$

The effect (coefficient) of initial shock, $\eta\beta\rho^{t-1}$, is positive but *monotonically decreasing* over time. Eventually, it dies out to zero. Initial wage differentials caused by uneven random luck among workers will not have a persistent significant effect on future wage differentials.

Case 3. $\rho=1$, and $\{\varepsilon_t\}$ is a stationary *AR (1)* process, i.e.

$$\varepsilon_t=\delta\varepsilon_{t-1}+u_t \text{ where } |\delta|<1. \quad (9)$$

We again assume no human capital depreciation in Equation (4) with $\rho=1$ as in Case 1, but the random component of wages, ε_t is assumed to be serially correlated following a stable AR(1) process. Equation (9) indicates that any kind of random luck in the labor market of an individual worker is correlated over time by δ. Provided that δ is positive, an individual worker is more likely to remain lucky in the next period once she gets lucky in the beginning. To the contrary, an unlucky worker is more likely to suffer from her continuous unluckiness in her later career. We may consider random luck as some unobserved individual heteroskedastic characteristic that helps a worker get a high-paying job or give a good signal to employers. Either directly or indirectly, these factors enable a worker to maintain her luck with employers and keep high wages over time. Since these factors are likely to be related with a worker's personality and attitude towards relationships, work, and risk, they produce a positive correlation with luck over time.

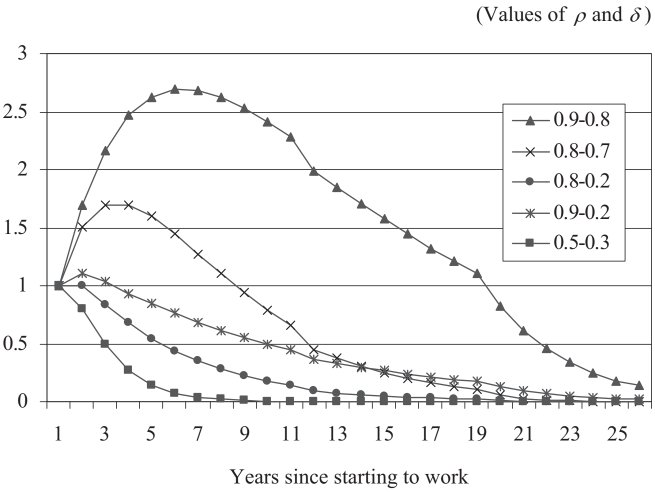
Further, the *AR(1)* assumption on the random component dynamics modifies the hours of work growth model in Equation (3) into the following:

$$h_t=h_0+\beta\delta^t\varepsilon_0+\beta\sum_{j=1}^{t}\delta^{t-j}u_j. \quad (10)$$

We predict that hours of work would evolve responding to the initial randomness to the extent of $\beta\delta^t$. If δ is equal (or close) to 1, then the effect of initial randomness on post working hours would be constant as β (or gradually diminish toward zero). If δ is close to -1, then the relation between initial randomness and post working hours would fluctuate between the upper bound β and the lower bound — β while converging to zero.

By applying Equation (4) to Equation (9), we obtain the following equation of human capital accumulation process which enables us to analyze the effect of temporally correlated random shock on future wages:

FIG. 1. SIMULATED CHANGES IN INITIAL WAGE EFFECTS (CASE 4) : $\eta\beta\left(\frac{\rho^t-\delta^t}{\rho-\delta}\right)$



$$\alpha_t=\alpha_0+t\eta\beta h_0+\eta\beta\left(\frac{1-\delta^t}{1-\delta}\right)\varepsilon_0+\eta\beta\sum_{j=1}^{t-1}\sum_{s=1}^{j}\delta^{j-s}u_s. \tag{11}$$

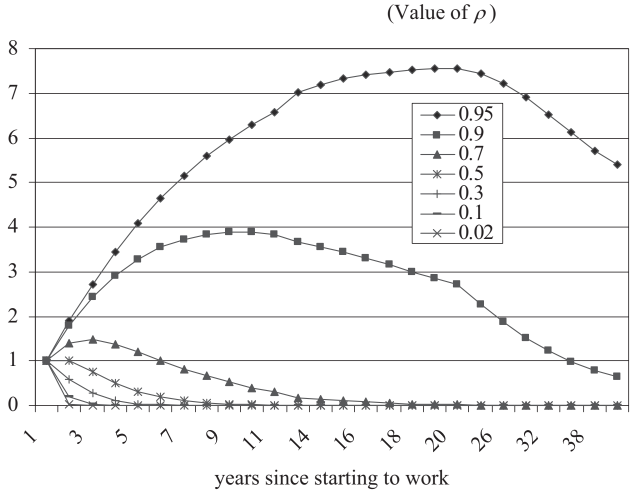
The initial shock produces a positive effect on future productivity and wage. This effect either *monotonically increases* over time with an upper bound of $\eta\beta\left(\frac{1}{1-\delta}\right)$ if δ is positive ($0<\delta<1$), or follows a fluctuating process around $\eta\beta\left(\frac{1}{1-\delta}\right)$ to which it eventually converges if δ is negative ($-1<\delta<0$) . In the case of positive δ , the positive and persistent impact of initial luck on future productivity and wages, given that other factors are fixed, renders the wage differential between lucky workers and unlucky workers in the first period at work stay persistent over time. This implies that the wage differential among experienced workers is a consequence of the variation of initial luck they receive when starting to work.

Case 4. Both $\{\alpha_t\}$ and $\{\varepsilon_t\}$ are stationary *AR (1)* processes. As our most general case, we combine human capital accumulation that allows for partial depreciation and an intertemporal correlation within random component in wages, i.e., α_t follows Equation (7) and ε_t follows Equation (9). Then it follows that:

$$\alpha_t=\rho^t\alpha_0+\frac{1-\rho^t}{1-\rho}\eta h_0+\eta\beta\left(\frac{\rho^t-\delta^t}{\rho-\delta}\right)\varepsilon_0+\eta\beta\rho^{t-1}\sum_{j=1}^{t-1}\sum_{s=1}^{j}\frac{\delta^{j-s}}{\rho^j}u_s. \tag{12}$$

Now, the marginal effect of initial shock on future wages, $\eta\beta\left(\frac{\rho^t-\delta^t}{\rho-\delta}\right)$ depends on two AR (1) parameters, ρ and δ. Assuming that both ρ and δ are positive numbers between 0 and 1, regardless of the relative magnitude of the two parameters, the effect of initial wage will eventually disappear. However, a converging trend of this effect depends on the specific choices

FIG. 2. SIMULATED CHANGES IN INITIAL WAGE EFFECTS (CASE 4, WHEN $\rho=\delta$): $\eta\beta(t\rho^{t-1})$



of two parameters. In particular, if $\rho=\delta$, then the effect of initial wage will be $\eta\beta(t\rho^{t-1})$.

Figure 1 illustrates simulated patterns of the initial random shock effect over time. Depending on different choices of ρ and δ, the marginal effect shows different trends. Even in cases of increasing trend, the trend is reversed to decrease at $t>6$. Intuitively, the reasonable conjecture is that the intertemporal correlation of a worker's productivity is stronger than that of her random shock in wages, say $\rho>\delta$. Figure 2 shows the simulated trend of initial shock when ρ is assumed to be equal to δ. If ρ is very close to 1, the marginal effect first explodes and then slowly decreases towards zero. As ρ gets smaller toward zero, the exploding trend stabilizes and the initial shock effect converges to zero within a relatively short period of time. Figures 1 and 2 uniformly show that changes in the marginal effect of initial shock on post wages are not necessarily monotonic. In Case 4, though the marginal effect of initial wages reduces to zero eventually, it does *not monotonically decrease* and the speed of convergence depends on the values of ρ and δ. Hence, during the early period of a working career, workers with positive initial shock might gain larger and larger benefits from it over certain time. Along with longer experience at work, it starts to decrease and eventually leaves no effect.

III. *Empirical Model*

We induce the prediction of wage growth by using Equations (1), (2) and (5) as follows:

$$w_t=\rho^{t-1}\alpha_0+\frac{1-\rho^{t-1}}{1-\rho}\eta h_0+\eta\beta\rho^{t-2}\varepsilon_0+\eta\beta\sum_{j=1}^{t-2}\rho^{t-2-j}\varepsilon_j+\varepsilon_t$$

Then, we apply four possible cases of the human capital accumulation process presented in Equations (6), (8), (10), and (11) to obtain the following wage equations:

Case 1. $$w_t=\alpha_0+\eta\beta\varepsilon_0+(t-1)\eta h_0+v_t \text{ where } v_t=\eta\beta\sum_{j=1}^{t-2}\varepsilon_j+\varepsilon_t, \tag{13}$$

Case 2. $$w_t=\rho^{t-1}\alpha_0+\eta\beta\rho^{t-2}\varepsilon_0+\frac{1-\rho^{t-1}}{1-\rho}\eta h_0+v_t \tag{14}$$

$$\text{where } v_t=\eta\beta\sum_{j=1}^{t-2}\rho^{t-1-j}\varepsilon_j+\varepsilon_t,$$

Case 3. $$w_t=\alpha_0+\eta\beta\left(\frac{1-\delta^{t-1}}{1-\delta}\right)\varepsilon_0+(t-1)\eta\beta h_0+v_t \tag{15}$$

$$\text{where } v_t=\eta\beta\sum_{j=1}^{t-2}\sum_{s=1}^{j}\delta^{j-s}u_s+\varepsilon_t, \text{ and}$$

Case 4. $$w_t=\rho^{t-1}\alpha_0+\eta\beta\left(\frac{\rho^{t-1}-\delta^{t-1}}{\rho-\delta}\right)\varepsilon_0+\frac{1-\rho^{t-1}}{1-\rho}\eta h_0+v_t, \tag{16}$$

$$\text{where } v_t=\eta\beta\rho^{t-2}\sum_{j=1}^{t-2}\sum_{s=1}^{j}\frac{\delta^{j-s}}{\rho^j}u_s+\varepsilon_t$$

In the empirical analysis, we use two different wage equation specifications to find out which cases more faithfully reflect the data generating process (i.e. the true effect of initial wages). Given a time period t, for each individual i, we consider two reduced form models of log post wage equations:

(Regression 1) $$w_t=\gamma w_0+X_t\delta_1+v_t,$$

and

(Regression 2) $$w_t=\gamma_0\alpha_0+\gamma_1\varepsilon_0+X_t\delta_2+v_t \quad ,$$

where w_0 is the logarithm of initial wage. X includes a set of explanatory variables to control for individual-specific demographic and labor market characteristics: age, years of schooling, AFQT score, gender, race, marital status, years of experience, and the squared years of experience. In Regression 2, since both initial ability (α_0) and random shock are unobserved, we need the first-step estimation of the initial wage equation given as:

$$w_0=Z_0\delta_0+\varepsilon_0, \tag{17}$$

where Z includes the set of observables by which we approximate α_0.We let Z satisfy the exclusion condition, that is, contain three groups of additional explanatory variables not included in X to avoid multicollinearity in Regression 2: first, individual-specific factors such as the squared years of education and an interaction term of years of education and AFQT score; second, the set of industry and occupation indicators as they are shown to generate variations in wages (Dickens, Katz, 1986); finally, the set of indicators for year of starting to work which would capture aggregate economic conditions and their effect on initial wage. Then, the initial wage effects are decomposed into a systematic part explained by Z and a random shock part. It is still possible that the predicted random component is correlated with unobserved individual innate ability since the comprehensive set of Z may not be exhaustive enough to consider all important systematic determinants of wages. To address this issue, we collect family information such as father's educational attainment, mother's educational attainment, the closest sibling's gender and educational attainment and use them as instruments in the parametric Two-Stage Least Squares estimation. The efficient nonparametric instrument is obtained by applying Newey's (1990) method to years of education and AFQT scores.

Using predicted initial wages and residuals from Equation (17) as $w_0 = Z_0\hat{\delta} + \hat{\varepsilon}_0 \equiv \hat{\alpha}_0 + \hat{\varepsilon}_0$, Regression 2 is estimated as follows:

$$w_t = \gamma_0\hat{\alpha}_0 + \gamma_1\hat{\varepsilon}_0 + X_t\delta_2 + v_t. \tag{18}$$

In Regression 1, we obtain the average total effect of the initial wage. Then, we decompose it into the effect of the observable systematic component, γ_0 and the effect of the unobservable random component, γ_1 in Regression 2. A series of questions we pursue here using the estimation of Equation (18) are (1) the presence of the effect of initial wage on future wage, (2) the persistence of this effect over time, and (3) the contribution of the initial random component in the effect of initial wage on post wages.

We further estimate the hours of work growth model we develop in Equations (3) and (10) using the logarithm of total annual hours of work in each post working year (1980 ~ 2000) and the logarithm of total annual hours of work in the initial year of work. For Equation (3), we approximate the 1-period lagged human capital using the 1-period lagged value of the predicted post wages. The predictions of post wages are implemented both without decomposition (Regression 1) and with decomposition (Regression 2) of initial wages.

IV. *Data and Sample*

For the empirical analysis described above, we employ the National Longitudinal Survey of Youth 79. Our original sample includes a cross-sectional individuals aged 14-21 as of December 31, 1978 (N=6,111). Based on the information about employment status over survey years 1979-2000, we indentify the year each respondent started to work and total years of experience. Then, we limit our sample to those who completed all schooling and have the year they started to work identified. The final sample includes 752 cross-sectional individual females.[2] Then initial wages are defined as the first wage observed *after* years of education completed stop increasing. This allows us to exclude cases where people start work during schooling or return to school after they start working.

For each year during 1980-2000, the current wage, our dependent variable, is the wage observed after 1 to 21 years later from the year of starting to work for each worker. For example, if the number of years since starting to work is equal to 3, the current wages are wages in 1982 for people who started to work in 1979, or wages in 1985 for those started to work in 1982, and so on. All wages are converted to the wages for compatibility using the

[2] Among these 752 individuals, we only consider those who have started to work no later than 1985. Then we are left with 716 individuals. All 36 individuals excluded from our final sample are high school graduates or less than high school educated. Unless some kind of interruption in schooling happens, these individuals are supposed to finish all schooling by age 18. If they have not started to work by 1985 (age 20-27), it implies that they spend 2-9 years without working. By restricting our sample to those who started working by 1985, we allow only limited years of job searching or after school training after finishing all education. In this way, we may exclude a possible discontinuous career effect on estimation results.

Table 1. Summary Statistics (in 1979)

Variable	Mean	Std. Dev.	Min.	Max.
Age	19.6	1.44	15	22
Years of schooling	11.7	1.39	6	16
AFQT scores	42.6	25.0	1	98
Being white	.87	.34	0	1
Being male	.47	.50	0	1
Being married	.25	.43	0	1
Year of experience (in 2000)	12.7	4.10	1	17
Father's education	11.0	3.00	0	20
Mother's education	11.1	2.34	0	18
The closest sibling's education	12.3	2.51	0	20
The closest sibling being male	.48	.50	0	1
The number of observations (N)	716			

Consumer Price Index. The distributions of years of starting work both overall and within a fixed number of years since the starting year are reported in Table A1. Table 1 presents summary statistics of the final sample individuals (N=716) used for the estimation.

V. *Estimation Results*

We implement a simple OLS estimation to find out which theoretical cases described in Section II are indeed consistent with the data. Results are presented in Tables 2 and 3. Table 2 and Figure 1 reports the results of Regression 1 where we estimate the effect of initial wages on future wages without decomposing initial wages, but controlling for age, educational attainments, AFQT score, gender, marital status, years of experience and its squared term. It is noteworthy that the effect of initial wage is persistent even after 20 years since workers started to work. On average, one percentage increase in initial wages raises all future wages by 0.288%. In this specification, we are unable to differentiate the effect of explained components of initial wage from that of random shock in initial wages.

Before we interpret this result as a *true* effect of initial wage, we pay attention to the possibility of an endogeneity problem. An endogeneity problem happens if any individual unobserved characteristic makes workers relatively more productive and raises both their initial wage and future wages. Innate ability is the common suspect considered responsible for this endogeneity bias. We are lucky to have a proxy variable for individual innate ability, AFQT scores measure in quantiles. By controlling AFQT scores explicitly in Regression 1, we try to avoid the possibility that the observed persistence of effects of initial wage over time reflects the persistent feature of workers' unobserved ability.

The second strategy to deal with endogeneity of initial wages is the instrumental variables estimation using educational attainment of family members as four instruments: father's education level, mother's education level, the closest sibling' education level completed in 1979 and the closest sibling's gender (being male). These four instrumental variables are used for the first stage estimation of initial wages to capture the endogeneity of initial wage variable in explaining post wage differentials. From the coefficient estimates in the first stage estimation, we decompose initial wages into systemically predicted part and unobserved random part. In

Table 2. Initial Wage Effects on Post Wages
(OLS estimates without decomposition)

Dependent Variable: Logarithm of Post Wage at t
(t denotes the number of years since starting to work)

t	Coeff. of w_0	t-Ratio	Std. Err.	No. Obs.	Adj. R^2
1	**.475**	13.1	.036	541	.374
2	**.500**	13.2	.038	521	.410
3	**.333**	6.62	.050	506	.226
4	**.211**	4.32	.049	498	.226
5	**.285**	6.03	.047	487	.295
6	**.381**	7.62	.050	498	.308
7	**.231**	4.97	.046	514	.288
8	**.186**	3.57	.052	514	.256
9	**.321**	4.93	.065	501	.192
10	**.248**	4.48	.055	508	.228
11	**.214**	4.59	.047	510	.318
12	**.266**	5.18	.051	488	.330
13	**.291**	3.94	.074	482	.225
14	**.235**	3.42	.069	484	.263
15	**.240**	4.08	.059	474	.274
17	**.271**	4.66	.058	455	.273
19	**.259**	4.41	.059	507	.252
21	**.244**	3.33	.073	447	.197

Notes: 1. Age, years of schooling, AFQT score, race/ethnicity (being white), gender (being male), marital status (being married), experience, and the squared experience are controlled for in obtaining the results reported above.
2. All coefficients are highlighted **bold** indicating their 1% statistical significance.

the second stage of estimation for post wage determination, we include the predicted initial wage and the first-stage residuals to separately examine the effects of the observed and the unobserved components of initial wages

As the correlation between these instruments and initial wage variable is rather weak (R-sq = 0.025), we explore the efficient IV estimation in nonlinear models using nonparametric least-squares cross-validation method (Li and Racine, 2007). The instrumental variables used for the nonparametric prediction of initial wages are AFQT scores and years of schooling as the family attributes are found little relevant for initial wage determination. We confirm the validity of the nonparametic IV by R-sq = 0.10, which supports the application of nonparametric IV to overcome the weak instrument problem of parametric IV.

Table 3 presents the results of the initial wage decomposition by three estimation methods which are explored to incorporate the possibility of endogeneity problem.[3] The OLS estimates of the unexplained random component of initial wages have positive and highly significant impacts on future wage persistently up to the most recent data available (year 2000). There is a slight fluctuation in the extent of the effects across time periods. The average effect of one percentage increase in initial wages is a 0.298% (OLS estimate) increase in future wages up to 21 years after having started work. This result provides supporting evidence for our theoretical hypothesis, which held that the positive random shock in initial wage increases future wages

[3] Results from the first-stage initial wage estimation are reported in Table A2. The full results of instrumental variables estimation for post wage determinations are available upon request to the corresponding author.

TABLE 3. DECOMPOSITION OF INITIAL WAGE EFFECTS

Dependent Variable: Logarithm of Post Wage at t
(t denotes the number of years since starting to work)

The effect of $\hat{\varepsilon}_0$	(1) OLS		(2) 2SLS		(3) Nonparametric IV	
t	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
1	**.442**	.041	**.433**	.043	**.466**	.037
2	**.500**	.042	**.490**	.043	**.506**	.038
3	**.349**	.056	**.335**	.058	**.329**	.051
4	**.235**	.054	**.210**	.056	**.214**	.049
5	**.285**	.053	**.284**	.054	**.285**	.048
6	**.440**	.056	**.443**	.058	**.376**	.050
7	**.243**	.052	**.236**	.052	**.243**	.047
8	**.184**	.058	**.166**	.060	**.195**	.053
9	**.316**	.075	**.230**	.061	**.329**	.066
10	**.266**	.063	**.258**	.064	**.259**	.056
11	**.222**	.052	**.201**	.053	**.211**	.048
12	**.277**	.057	**.261**	.058	**.271**	.052
13	**.346**	.081	**.363**	.084	**.283**	.074
14	**.220**	.075	**.212**	.077	**.234**	.069
15	**.254**	.063	**.245**	.066	**.232**	.059
17	**.290**	.062	**.273**	.064	**.282**	.059
19	**.251**	.063	**.241**	.065	**.270**	.059
21	**.235**	.077	**.218**	.079	**.247**	.074

Notes: 1. The predicted initial wages and residuals are instrumented by four excluded instruments (fathers' education level, mother's education level, the closest sibling's education level and the closest sibling's gender as being male) and a set of included instruments such as age, education, education squared, AFQT score, an interaction term of education and AFQT score, race, sex, marital status and the sets of industry indicators, occupation dummies and indicators for the year of starting to work.

2. The dependent variable is the logarithm of post wages at t where t denotes the number of years since the year of starting to work.

3. Control variables include the predicted initial wage, age, years of schooling, AFQT score, race/ethnicity (being white), gender (being male), marital status (being married), years of experience, and the squared experience.

4. All coefficients of the predicted initial wage variable are found statistically insignificant at the 10% level. Estimates are available upon request to authors.

5. All estimated coefficients for $\hat{\varepsilon}_0$ are highlighted **bold** indicating that they are significant at the 1% level.

6. Newey nonparametric efficient instrument for the initial wage is estimated using years of education and AFQT scores in the platform of n-program developed by J. Racine.

permanently through a worker's human capital accumulation. Figure 4 plots the estimates of initial random component effects. Certainly, there are fluctuations in magnitude over time. For the first 3 years, the coefficients are remarkable higher than for the later years. Whatever the causal relationship between initial wages and post wages may be, it is reasonable to assume that initial wages should have a strong relationship with post wages over a short time period. Over a 3-9 year period, the effect of initial random shocks fluctuates across years with a slightly decreasing trend. Over time, the extent of the fluctuation in the initial random shock effect stabilizes. After 10 years, it settles down to around 0.24 and fails to converge to zero. It is noteworthy that although there is modest variation in values of estimates among three approaches, the estimated effect of initial randomness seems remarkably consistent across different estimation methods and persistent over time.

Fig. 3. Initial Wage Effects (OLS Estimates without Decomposition)

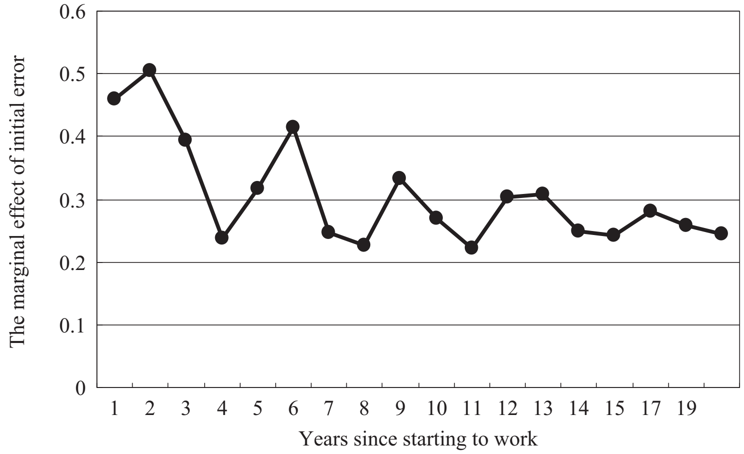
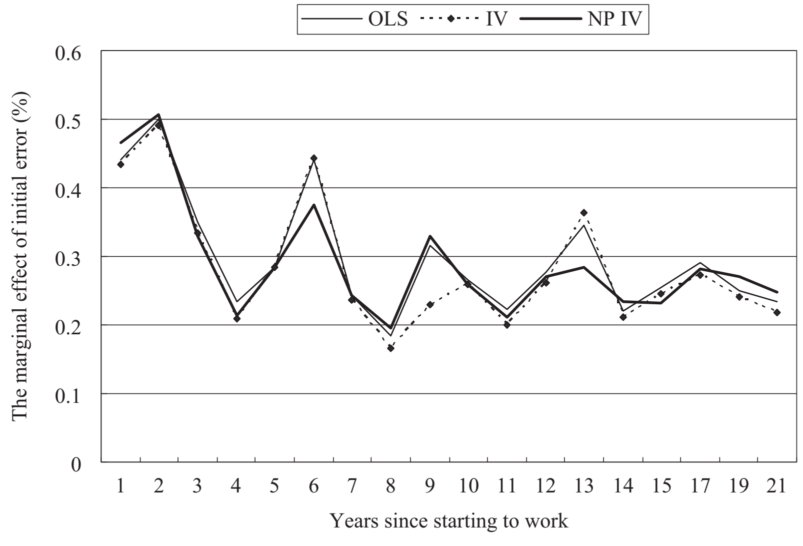


Fig. 4. Estimates of The Effect of Initial Randomness (by decomposition)



It is important to integrate our empirical results with one of the theoretical predictions. As shown in Figure 4, the effect of initial randomness is decreasing during only a short period time and it does not disappear toward zero even 21 years later. This seems to support more the prediction of Case 1 than three other Cases. Given that our sample only contains relatively young workers (aged 15 to 22 in 1979, and aged 36 to 43 in 2000), the no human capital depreciation assumption seems not a restrictive assumption. Our empirical results support the proposition that if a worker's chance in unobservable factors in wage offer is purely random due to insufficient information for the employers regarding her true productivity, it functions as an incentive for her to spend more time in working and to invest more effort in work. This leads

TABLE 4. HOURS OF WORK GROWTH

Dependent Variable: Log(Total Annual Hours of Work at t) (t denotes the number of years since starting to work)								
The effect of	W/O decomposition		With decomposition					
$w_t - \alpha_{t-1}$			(1) OLS		(2) 2SLS		(3) Nonparametric IV	
t	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
2	**.123**	.056	**.129**	.057	**.136**	.059	**.121**	.057
3	.119*	.068	.121*	.068	.105*	.071	.112*	.068
4	**.221**	.059	**.225**	.060	**.214**	.062	**.221**	.059
5	**.117**	.058	**.115**	.059	.106*	.061	**.114***	.059
6	**.207**	.075	**.203**	.075	**.211**	.079	**.208**	.075
7	**.172**	.063	**.176**	.064	**.147**	.064	**.172**	.063
8	**.346**	.058	**.346**	.058	**.302**	.060	**.355**	.059
9	**.133**	.054	**.134**	.054	**.154**	.056	**.136**	.054
10	.051	.056	.052	.056	.050	.059	.050	.056
11	.045	.045	.044	.045	.030	.046	.044	.045
12	.091*	.051	.090*	.051	.084	.053	.088*	.052
13	**.189**	.051	**.187**	.051	**.176**	.053	**.190**	.051
14	**.125**	.045	**.119**	.046	**.109**	.047	**.125**	.045
15	.089*	.049	.089*	.049	.077	.050	.089*	.049
17	.009	.056	.009	.056	.0005	.049	.009	.056
19	.002	.057	.009	.057	-.016	.058	-.002	.057
21	**.127**	.051	**.121**	.050	.089*	.050	**.122**	.050

Notes: 1. Results are obtained from the hours of work growth model in Equation (3). In all estimations, we control for the logarithm of the initial hours of worked.
2. α_{t-1} is predicted as the 1-period lagged value of the predicted post wages. In predicting post wages, we apply four different estimation methods: first, without decomposition of the initial wage (refer to Regression 1), secondly, with decomposition of the initial wage into the systematic factor and the random factor using OLS, 2SLS and the nonparametric IV methods (refer to Regression 2).
3. As we use the 1-period lagged post wages for the proxy of α_{t-1}, estimates at $t=1$ is not available.
4. The highlighted coefficients in **bold** are statistically significant at the 5% level. * indicates the 10% statistical significance.

to a permanent increase in her human capital level. Therefore, after over 20 years later, her post wages benefit from the good draw in her initial wage.

An interesting explanation of wage differential within groups is provided. Given the same innate ability, educational attainment and other demographic conditions, workers benefit or suffer from wage inequality, which is caused by differences in their initial positive draw at work. As the initial randomness in wages influences how individual workers behave at work, the observed persistent wage differential will not be eliminated simply by equalizing systematic determinants of wages across workers.

Tables 4 and 5 further present the estimation results of the hours of work growth model in Equations (3) and (10), respectively. In Table 4, we apply four different approaches in predicting α_{t-1} and find evidence for the positive response of labor supply to the random shock in wages, which supports our prior postulation on the sign of β in Equation (3). Findings in Table 5 further show that labor supply of an individual is positively related with the initial randomness in wages. We plot the statistically significant coefficients of the initial randomness

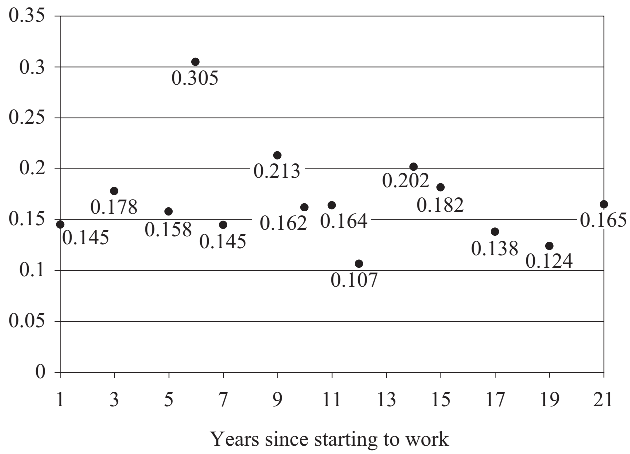
TABLE 5. INITIAL WAGE EFFECT ON HOURS OF WORK

Dependent Variable: Log(Total Annual Hours of Work at *t*)
(*t* denotes the number of years since starting to work)

Variable	$\text{Log}(h_0)$		$\hat{\varepsilon}_0$	
t	Coeff.	Std. Err.	Coeff.	Std. Err.
1	**.371**	.031	**.145**	.060
2	**.306**	.034	.099	.065
3	**.225**	.038	**.178**	.074
4	**.088**	.038	.101	.073
5	**.187**	.044	**.158**	.081
6	**.143**	.047	.305*	.087
7	**.136**	.039	**.145**	.073
8	**.117**	.035	.095	.064
9	**.151**	.038	**.213**	.070
10	**.131**	.036	**.162**	.069
11	.040	.032	**.164**	.059
12	-.008	.032	.107*	.059
13	**.108**	.039	-.018	.076
14	**.089**	.040	**.202**	.077
15	.040	.035	**.182**	.068
17	**.120**	.039	.138*	.077
19	**.083**	.036	.124*	.071
21	.004	.032	**.165**	.060

Notes: 1. Results are obtained from the estimation of Equation (10): the hours of work growth model with the AR (1) process of the randomness in wages.
2. $\hat{\varepsilon}_0$ are obtained from the Newey nonparametric efficient instrumental variable (IV) estimation of the initial wage equation.
3. The highlighted coefficients in **bold** are statistically significant at the 5% level. * indicates the 10% statistical significance.

FIG. 5. THE EFFECT OF INITIAL RANDOMNESS ON HOURS OF WORK



in Figure 5 and note two features: First, the randomness in initial wages has a positive effect on post labor supply, which is persistent up to 21 years since the year of starting to work. Secondly, this effect shows a modest deceasing pattern over time.

VI. *Conclusion*

The standard Mincer wage regression developed by Mincer (1974) and its numerous successors leave ambiguity in explaining what are unobserved factors in wage determination and what causes wage differentials across individuals. Among all possible unobserved factors, we examine the role of initial wage in understanding the determination of future wages. We present a human capital accumulation model showing the theoretical predictions of initial wage effects on future wages. We estimate the future wage equation as a function of initial wage which is decomposed as a systemically explained part and an unexplained random part. A positive persistent effect of initial randomness is predicted by the human capital model and supported by the estimation results. This suggests that the level of initial wage, specifically its random portion rather than the actual productivity portion related to observables, provides an additional incentive for workers to invest more effort at work and thus to maintain their wages high throughout their working life.

Some issues warrant further research. One may argue that the random component of initial wage is related to an individual's ability not observed by economists but partially observed by firms who make a hiring. Then, the initial wage effects may reflect unobserved individual heterogeneity rather than randomness by information incompletion. We address this issue using the parametric and nonparametric efficient instruments and find the conclusions from the OLS estimates are largely robust. Yet, it is worthwhile to explore whether there are strong instruments based on information about family members such as siblings' initial wages and their AFQT score, which we believe to serve as valid alternative instruments. Unfortunately, we are unable to find these variables in the present NLSY data.

Another question is the mechanism in which the randomness in initial wage is created. Obviously, educational attainment, experience, and other well-accepted determinants of wages do not account for it. Firm's limited capability in assessing the worker's true productivity may be the reason for the randomness in wage offer, making the wage unequal to the worker's contribution to the firm. Further studies need to pay attention to the role of employers' behavior in generating the disparity between the actual wage and the worker's productivity.

References

Baker, M. (1997), "Growth-Rate Heterogeneity and the Covariance Structure of Life-Cycle Earnings", *Journal of Labor Economics* 15(2), pp.338-375.

Becker, G. (1975), *Human Capital*, Chicago: University of Chicago Press.

Ben-Porath, Y. (1976), "The Production of Human Capital and the Life Cycle of Earnings", *Journal of Political Economy* 75, pp.352-365.

Bertrand, M., S. Mullainathan (2001), "Are CEOs Rewarded for Luck? The Ones without Principals Are", *The Quarterly Journal of Economics* 116(3), pp.901-932.

Cossa, R. D., J. Heckman, and L. Lochner (1999), *Wage Subsidies and Skill Formation: A study of the EITC*. E. Phelps (Eds.), Subsidies to Low-Wage Employment, New York: Russell Sage Foundation.

Heckman, J., L. Lochner, and R. D. Cossa (2002), "Learning-By-Doing vs. On-The-Job Training: Using Variation Induced by The EITC to Distinguish Between Models of Skill Formation", *NBER Working Paper* No. 9083, Cambridge: MA.

Dickens, W. T., L. F. Katz (1986), "Interindustry Wage Differences and Industry Characteristics", *NBER Working Paper* No. 2014, Cambridge: MA.

Garvey G. T., T. Milbourn (2006), "Asymmetric Benchmarking in Compensation: Executives are Rewarded for Good Luck but Not Penalized for Bad", *Journal of Financial Economics* 82, pp.197-225.

Li, Q., J. S. Racine (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton University Press: New Jersey.

Mincer, J. (1974), *Schooling, Experience and Earning*, New York: Columbia University Press.

Neumark, D. and P. Taubman (1995), "Why Do Wage Profiles Slope Upward? Tests of the General Human Capital Model", *Journal of Labor Economics* 13(4) pp.736-761.

Newey W. K. (1990), "Efficient Instrumental Variables Estimation of Nonlinear Models", *Econometrica* 58(4), pp.809-837.

Appendix

TABLE A1. SAMPLE COMPOSITION BY YEAR OF STARTING WORK

Years Since	Year of Starting Work							
Initial Work	79	80	81	82	83	84	85	Sample Size
1	451	68	19	6	6	9	4	563
2	438	60	18	6	8	9	6	545
3	421	60	19	6	9	9	8	532
4	407	61	20	9	9	10	7	523
5	421	58	17	5	6	10	6	523
6	411	65	19	7	4	9	6	521
7	428	69	23	6	3	9	7	545
8	433	63	20	5	3	11	6	541
9	419	64	21	7	3	11	6	531
10	417	70	24	7	4	9	7	538
11	427	66	22	6	6	10	7	544
12	406	68	25	8	5	12	7	531
13	409	67	25	8	5	13	0	527
14	423	67	25	8	4	0	0	527
15	438	66	27	10	0	0	0	541
17	450	76	29	0	0	0	0	555
19	466	71	0	0	0	0	0	537
21	476	0	0	0	0	0	0	476

Note: 1. The sample size for x years since initial work corresponds to estimation results reported in Table 3, with x ranging 1 to 21.

2. For estimation of x years since initial work, the dependent variable is logarithm of post wage at x + the year of starting to work. For example, wages in 1982 for workers who have started to work in 1979 and wages in 1983 for workers who have started to work in 1980 and so on are used in the estimation for 3 years since initial work in Table3.

Table A2. Initial Wage Regression

Dependent Variable: Initial Wage, w_0			
Control Variable	Coefficient	t-Ratio	Std. Err.
Constant	2.974***	4.57	.6502
Age	.047***	4.22	.0120
Education	.261**	2.16	.1288
Education^2	-.013**	-2.00	.0067
AFQT	-.005	-.66	.0079
AFQT* Education	-.0007	.98	.0007
White	-.012	-.64	.0481
Male	.205***	7.92	.0314
Married	-.037	-.87	.0368
No. Obs.		697	
Adjusted R^2		.337	

Notes: 1. Start 79-85 dummies for the year of starting to work between 1979 and 1985.
2. The reported result is obtained from the first step estimation of initial wage associated with Table 3.
3. Dummies for the year of starting to work (1979～1985), industry indicators (for 7 categories) and occupational indicators (for 3 categories) are included.