The Impact of Major Job Mismatch on College Graduates' Early Career Earnings

NILS Working Paper Series
Working Paper No. 170/2011

Rong Zhu

National Institute of Labour Studies
Flinders University, Adelaide, Australia

The Impact of Major–Job Mismatch on College Graduates' Early Career Earnings*

Rong Zhu†

Abstract

In this paper, I assess the impact of the mismatch between college major and job on college graduates' early career earnings using a sample from China. I find that on average a major–job mismatched college graduate suffers from a small income loss. I argue that Chinese Universities' emphasis on both general skills and major-specific skills could possibly explain why the average income penalty from major–job mismatch is very limited for college graduates in China. I also find that the income loss is heterogeneous and skewed that about one third of the major–job mismatched college graduates earn more than those matched ones.

JEL classification: I2; J31

Keywords: Major–job mismatch; college graduates; local linear kernel estimation

1 Introduction

Recent decades have witnessed a growing literature on the examination of the effects of education–occupation mismatch on labor market outcomes. According to the assignment theory by Sattinger (1993), returns to human capital investment could vary with the quality

*I would like to thank two anonymous referees for insightful comments. I am also grateful to Adrian Pagan, Denzil Fiebig, Elisabetta Magnani, Shiko Maruyama, Yuanyuan Gu, seminar participants at the University of New South Wales and conference participants at the 4th International Symposium on Contemporary Labor Economics at Xiamen University for helpful suggestions. Data provided by the China Survey Data Network for this project is also gratefully acknowledged. The usual disclaimer applies.

†National Institute of Labour Studies, Faculty of Social and Behavioural Sciences, Flinders University, Adelaide 5001, Australia. E-mail address: rong.zhu@flinders.edu.au. Phone: +61412726566.

of the match between worker and job. When the attained schooling of an individual is different from the years of schooling required for his/her job, measures of overeducation or undereducation could be constructed to study the labor market effects of this quantity–based education–occupation mismatch (Hartog, 2000; McGuinness, 2006).

Previous studies have primarily focused on the wage effects of overeducation. Most of the existing literature find that overeducated workers earn less than equally educated ones whose jobs match their years of education. The estimate of overeducation wage penalty ranges from 7.4% to 27%, varying with the data sets and measures of overeducation used (McGuinness, 2006). Other empirical studies center on the effects of overeducation on career mobility (McGoldrick and Robst, 1996; Buchel and Mertens, 2004), cognitive decline (De Grip et al., 2008), job satisfaction (Allen and van der Velden, 2001) and migration decisions (Quinn and Rubb, 2005).

The quantity of schooling is not the sole way to consider the match between a worker's education and job. The mismatch between one's field of college education and job has been ignored for a long time, and it has begun to capture researchers' attention in recent years. With cross sectional data from 13 European countries, Wolbers (2003) explores the effects of major–job mismatch on the labor market positions of school–leavers. He finds that school–leavers with a non–matching job generally achieve lower occupation status than those with a matching job. He also finds that school–leavers with mismatched jobs are more likely to engage in on–the–job search and have a higher probability of vocational training participation. Regarding the wage effects of major–job mismatch, the two existing studies I know of show the importance of full utilization of major–specific human capital in jobs. Using data derived from the 1993 US National Survey of College Graduates, Robst (2007a) discovers that major–job mismatch is associated with about 11% loss of annual earnings. With data from Sweden, Nordin et al. (2010) find that major–job mismatch could lead to a 20% decrease in annual earnings for male college graduates. The corresponding income penalty for female graduates is about 12%.

While the aforementioned empirical studies provide important findings of the effects of major–job mismatch on college graduates' labor market outcomes, there is some void in the literature. First, whether and how this particular type of occupation–education mismatch affects college graduates' earnings in a developing country is still unavailable. It is likely that the previous findings do not hold for college graduates in low and middle income countries. One possible reason for the potential difference in the income penalties of major–job mismatch between a developed country and a developing country lies in the differences in tertiary education system. Usually, college education is regarded as a bundle of general and occupation–specific skill education (Robst, 2007b). General skills are believed to be transferrable between different occupations, and they can be rewarded in all occupations. However, occupation–specific skills are expected to be remunerated only within major matched occupations. As a result, the differential emphasis on general and occupation–specific skills in college education in different countries could result in different wage effects. For example, Nordin et al. (2010) find substantial income penalties from major–job mismatch for Swedish college graduates. Their explanation for this finding is that most fields of the higher education in Sweden are very specialized relative to many other countries and college graduates learn more occupation–specific skills and relatively fewer general skills. As discussed later in this paper, both general and specific skill education have been emphasized in Chinese universities. The wage penalty from major–job mismatch for Chinese college graduates could be lower than the penalty found in Sweden. Second, the existing literature relies heavily on OLS regression to estimate the mean wage effects of major–job mismatch. However, if the commonly–used linear functional form assumption is not an adequate description of data generating process, a misspecified model could lead to inconsistent estimates and misleading policy prescriptions.

In this paper, using a data set on graduates from the colleges located in China's Shandong province, I assess the impact of major–job mismatch on college graduates' early career earnings. While China has experienced robust economic growth in recent decades, it has also experienced an unprecedented expansion in higher education institute enrolment since the late 1990s. During the two decades from 1978 to 1998, the tertiary student enrolment in China had gradually increased from 0.86 to 1.08 million (Bai, 2006). As a response to the growing demand for highly-qualified manpower under the context of economic globalization and continued economic reform, the Chinese government began to accelerate its pace of tertiary education growth towards mass higher education in 1999.[1] From 1999 to 2004, the growth rates in college enrolment of freshmen were respectively 44.8%, 31.5%, 21.3%, 19.8%, 19.0% and 17.0%. The corresponding college freshman numbers had increased from 1.16 million in 1998 to 4.47 million in 2004.[2] With such a large number of college graduates, it is almost impossible for the labor market to allocate college graduates to the jobs that perfectly match their fields of college education. Actually, Zhao and Sheng (2008) find that the rapid growth in higher education is not consistent across majors, which may result in more imbalances between supply and demand among college graduates in certain majors. As a result, examining the labour market effects of major-job mismatch in China is both of interest and importance. As my sample is based on the college graduates who have left college for six to 12 months, the analysis of the wage effects of a major–job mismatch not only contributes to the education–occupation mismatch literature, but also contributes to the school–to–work transition literature (Ryan, 2001) and the literature on the youth labor market problem (Freeman and Wise, 1982).

I do the estimation using the local linear kernel estimation developed by Li and Racine (2004), a nonparametric method that can smooth both continuous and categorical variables. Unlike most parametric models, such as OLS regression, this technique does not impose any functional form or distribution assumptions. As a result, the nonparametric estimates I have obtained are more robust. Different from most nonparametric techniques requiring variables to be continuous, this method is especially appropriate for me given the categorical predictors I have. A further advantage of this method is that it could estimate the wage effect of major–job mismatch for every mismatched college graduate, which allows me to check the heterogeneity in the treatment effects and summarize the effects for different subgroups.

I find that on average, major–job mismatched individuals only suffer from a monthly income loss of 1.28%, much lower than the 5.90% estimated using OLS regression. The

[1]Interested readers are referred to Chen (2005) and Bai (2006) for more background about the China's higher education reform.

[2]Sources: Yearbook of China's Educational Statistics (2000); Surveys of Educational Reform and Development in China (2001, 2002, 2003, 2004).

specification test by Hsiao et al. (2007) rejects the correctness of my linear specification at conventional confidence levels, showing that the OLS estimates I have obtained have exaggerated the wage effects of major–job mismatch among early career college graduates. I argue that Chinese universities' emphasis on both general and specific skills could possibly explain why the average penalty is very limited for college graduates in China.

With individual–specific estimates, I find that about one third of the Chinese college graduates in my sample benefit from major–job mismatch in terms of earnings. This finding has not been documented before, but it is consistent with the discussion in Robst (2007b). Robst (2007b) has argued that workers accepting a mismatched position due to supply–related factors such as career oriented reasons (pay or promotion opportunities) may earn higher wages than well matched workers. I also find that the income losses are heterogeneous. Among the human capital theory, job competition theory and the assignment theory that are related to the phenomenon of education–occupation mismatch in labor markets, my finding that wage levels are dependent on college education, job characteristics and the match between major and job provides support for the assignment theory only.

The remainder of the paper is organized as follows. Section 2 presents the empirical approach taken. The following section is a description of the data set and variables. In Section 4, I discuss the sorting–by–ability problem that may affect the reliability of my estimates. In Section 5, I present and discuss the regression results for the full sample and different subsamples. Section 6 concludes.

2 Empirical Methodology

I resort to nonparametric local linear kernel estimation developed by Li and Racine (2004) to do the estimation. As mentioned in the Introduction, I choose this technique for three reasons: (i) The widely–used linear regression relies on functional form assumptions. When misspecified, the linear specification could lead to inconsistent estimates and misleading policy prescriptions. The nonparametric approach does not impose any functional form or distribution assumptions, and it allows for any kind of interactions among all variables. As a result, the nonparametric method is more robust; (ii) Linear regression only provides me with a mean estimate, which might conceal how heterogeneous the effects could be. With the local linear approach, I could obtain an observation–specific coefficient estimate for each individual, and this would allow me to explore the heterogeneity in coefficient estimates; (iii) As shown later in the data description section, I have many discrete control variables for this analysis. Most nonparametric methods require the explanatory variables to be continuous. The local linear method by Li and Racine (2004) could smooth both continuous and categorical variables, and this allows me to do this applied work under minimum assumptions.

The model is specified as: $W_i{=}F\left(X_i, Y_i, Z_i\right)+\varepsilon_i$, where $F\left(X_i, Y_i, Z_i\right)$ is the conditional mean function to be estimated. $X_i{=}(x_{1i}, x_{2i}, \ldots, x_{si})$ is a vector of continuous variables and s is the dimension of X_i. $Y_i{=}(y_{1i}, y_{2i}, \ldots, y_{li})$ is a vector of unordered categorical variables and l is its dimension. $Z_i{=}(z_{1i}, z_{2i}, \ldots, z_{ti})$ is a vector of ordered discrete variables and t denotes the dimension of Z_i.

I find $\gamma\left(X_j\right){=}(\alpha_j, \beta\left(X_j\right))'$ to minimize the following weighted least squares problem:

$$\sum_{i=1}^{N}\left(W_i-\alpha_j-\left(X_i-X_j\right)'\beta\left(X_j\right)\right)^2 K(\widehat{h}, \widehat{\lambda}, \widehat{\mu}) \tag{1}$$

where α_j predicts W_j; $\beta\left(X_j\right)$ is the vector of the partial derivative of $F\left(X_j, Y_j, Z_j\right)$ with respect to X for individual j; $K(\widehat{h}, \widehat{\lambda}, \widehat{\mu})$, the multivariate product kernel for mixed data types, is equal to $\prod_{q=1}^{s}\frac{1}{\widehat{h}_q}g(\frac{x_{qi}-x_{qj}}{\widehat{h}_q})\prod_{q=1}^{l}m(y_{qi}, y_{qj}, \widehat{\lambda}_q)\prod_{q=1}^{t}n(z_{qi}, z_{qj}, \widehat{\mu}_q)$. In the generalized product kernel, g is the second–order Gaussian kernel and h_q is the bandwidth for the qth component of X. m is the kernel function for an unordered discrete variable. It is a variation of Aitchison and Aitken (1976) kernel function, which equals to 1 if $y_{qi}{=}y_{qj}$ and λ_q $(0{\leq}\lambda_q{\leq}1)$ otherwise. For an ordered categorical variable, I use the kernel function $n(z_{qi}, z_{qj}, \mu_q)$ proposed by Li and Racine (2004), which equals 1 if $z_{qi}{=}z_{qj}$ and $\mu_q^{|z_{qi}-z_{qj}|}$ $(0{\leq}\mu_q{\leq}1)$ otherwise. It should be noted that the bandwidths are allowed to differ across variables. It is easy to see that if all the smoothing parameters selected for discrete variables are equal to 0, the product kernel $\prod_{q=1}^{l} m(y_{qi}, y_{qj}, \widehat{\lambda}_q) \prod_{q=1}^{t} n(z_{qi}, z_{qj}, \widehat{\mu}_q)$ for discrete variables becomes an indicator function. In this sense, the conventional frequency–based kernel method is a special case of the nonparametric approach I use.

After running the OLS regression of W_i on $(1,\, X_i - X_j)$ with weight $[K(\widehat{h}, \widehat{\lambda}, \widehat{\mu})]^{\frac{1}{2}}$, the consistent estimates could be obtained as:

$$\widehat{\gamma}(X_j) = \left(\widehat{\alpha}_j, \widehat{\beta}\,(X_j)\right)' = \left[\sum_{i=1}^{N} K(\widehat{h}, \widehat{\lambda}, \widehat{\mu}) \begin{pmatrix} 1 & (X_i - X_j)' \\ X_i - X_j & (X_i - X_j)(X_i - X_j)' \end{pmatrix}\right]^{-1} \sum_{i=1}^{N} K(\widehat{h}, \widehat{\lambda}, \widehat{\mu}) \begin{pmatrix} 1 \\ X_i - X_j \end{pmatrix} W_i \tag{2}$$

To estimate the coefficients of categorical variables, I follow the counterfactual analysis adopted in Henderson et al. (2006). In my analysis, the coefficient for the binary variable *Mismatch* is calculated as the counterfactual change in the dependent variable if the mismatched individual switches from *Mismatch*=0 to *Mismatch*=1, other variables being the same. For those individuals with *Mismatch*=0, the effect is 0. Suppose mismatch is the kth unordered variable, then the coefficient estimate of y_{kj} for the mismatched individual j is calculated as:

$$\widehat{\delta}(y_{kj} = 1) = \widehat{F}(X_j, y_{1j}, \ldots, y_{kj} = 1, \ldots, y_{lj}, Z_j) - \widehat{F}(X_j, y_{1j}, \ldots, y_{kj} = 0, \ldots, y_{lj}, Z_j) \tag{3}$$

For a discrete variable with more than two categories, a reference category is specified. Similarly, the coefficient estimate is calculated as the counterfactual change in the dependent variable if the individual switches from the reference category to the factual category, *ceteris paribus*. Note that both the coefficients of categorical variables and continuous variables are observation–specific. Without integration over other control variables, each coefficient estimate obtained is a conditional estimate.

I employ the widely–used Least Square Cross Validation (LSCV) procedure to select the bandwidths (h, λ, μ) (Pagan and Ullah, 1999; Li and Racine, 2007). It is based on minimizing the integrated square errors between the estimated distribution and the true distribution. Li and Racine (2004) shows that the bandwidths selected using this data–driven procedure are asymptotically optimal in the sense of minimizing the asymptotic mean square errors. The objective function is given by:

$$CV(h, \lambda, \mu) = \frac{1}{N} \sum_{i=1}^{N} (W_i - \widehat{F}_{-i}(X_i, Y_i, Z_i))^2 \tag{4}$$

where $\widehat{F}_{-i}\left(X_i, Y_i, Z_i\right)$ is the leave–one–out estimator of $F\left(X_i, Y_i, Z_i\right)$ and (h, λ, μ) are the smoothing parameters selected to minimize the objective function. I refer to Li and Racine (2004, 2007) for more details about the kernel method and Hayfield and Racine (2008) for the implementation of this approach. My empirical approach here is similar to that in Henderson et al. (2006). For an example of the application of this nonparametric approach education economics, I refer to Eren and Henderson (2008), which examines the differential impacts of homework on student achievement.

3 Data

3.1 Data and Variables

I use the data from the 2008 Chinese College Graduates' Employment and Work Skills Survey for this analysis. The survey is a national one which covers college graduates who share similar years of schooling and have been away from college for about six to 12 months. Respondents were asked questions pertaining to individual characteristics, college education and job information.

I only have access to the 2008 data set on the workers who graduated in 2007 from four–year colleges located in the Shandong province. These workers were randomly sampled and they all have fulltime jobs. After dropping observations with missing variable information, the final sample I obtained consisted of 5879 observations. In the final sample, individuals had graduated from 43 colleges in Shandong province, including four "211 Project Universities".[3] These workers were from 268 majors and 96 industries. Guided by the "2004

[3]The "211–Project Universities" consist of about 100 top research universities that account for 6% of the 1700 institutions of tertiary education in China. The figure of 21 and 1 within 211 represent the 21st

Major Catalogues for General Higher Education Institutions", issued by the Ministry of Education of China, I group these majors into 10 categories, namely, Natural Sciences, Engineering, Agriculture, Medicine, Economics, Management, Law, Education, Literature as well as History and Philosophy. I also group those observations into eight industries: Manufacturing, Construction, Financial Service, Education, Information, Business Service, Logistics and Transportation, and Other Industries. These groupings will help understand how the wage effects of major–job mismatch could differ by major and industry categories.

The dependent variable $Log(Income)$ in my following regression analysis is the natural logarithm of monthly income that a college graduate received from employer in 2007.[4] The binary variable $Mismatch$ is equal to 1 if the individual had a major-mismatched job and 0 otherwise. This variable is created from respondents' answers to the question about their current job status. The two choices are given as *"I am now employed, and the job is related to my major"* and *"I am now employed, but the job is unrelated to my major"*. I consider a graduate who chose the second answer as a worker with a job completely mismatched to his/her college major. I regard the first answer as the case that there was a major–job match for the respondent, although it could be closely or weakly matched.[5] In the following analysis, a binary variable indicating a complete mismatch, or not, is used.[6] While this measure of a major–job mismatch is subjective, it is common to see subjective measures of required schooling used in the overeducation literature (Sicherman, 1991; Dolton and Vignoles, 2000; Allen and van der Velden, 2001; Korpi and Tahlin, 2009). As argued by Robst (2007a), in the case of major–job mismatch, a college graduate is expected to have

century and approximately 100 universities respectively. These four "211–Project Universities", located in the Shandong Province, are Shandong University, The Ocean University of China, The China University of Petroleum (Eastern China) and Harbin Institute of Technology (Weihai).

[4]Ideally, I would like to examine hourly wages. However, the data does not contain any information on working hours.

[5]It is tempting to distinguish between close match and partial match. However, the data limitation prevents me from doing this. In the existing literature, Wolbers (2003) also does not consider the difference between close match and weak match. Both Robst (2007a) and Nordin et al. (2010) consider the differences in the extent of match, but their estimation results all show that in terms of wage effects, weakly matched individuals are much more comparable to matched individuals than to those completely mismatched.

[6]It is possible that major–job mismatch and overeducation/undereducation both exist among these college graduates. However, the data does not contain enough information for me to determine whether a college graduate is undereducated or overeducated. In this paper, I only focus on the wage effects of major–job mismatch. Interested readers are referred to Robst (2008), who considers the wage effects of both the horizontal and vertical education–occupation mismatch.

less incentive to misreport than in the case of reporting on the years of schooling, and it is difficult to develop an objective algorithm for determining whether a major and a job are unrelated or not.

A set of variables representing the characteristics of individuals, attainment of education, jobs and regions are included as covariates in the following estimation. Individual characteristics include *Age* and *Gender*; education variables include major category (*Major*) and an indicator variable showing whether the university is one of the "211–Project Universities" in China (*Keyuniv*); job's characteristics include the ownership and the employment size of the work unit (*Ownership*, *Firmsize*), as well as current job tenure (*Tenure*); region characteristics consist of the types of the city and the province where the employee works (*Citytype*, *Provtype*), a binary variable indicating whether one works within his/her home province (*Homeprov*), or Shandong province (*Collprov*) and a binary variable showing whether the college graduate works in the city where the college is located (*Collcity*). Corresponding to the notation in Section 2, W is the logarithmic monthly income, X=(*Age*, *Tenure*), Y=(*Mismatch*, *Gender*, *Homeprov*, *Collprov*, *Collcity*, *Major*, *Ownership*, *Provtype*, *Industry*), and Z=(*Keyuniv*, *Firmsize*, *Citytype*). See Table 1 for the detailed definitions and measurements of these variables.

3.2 Descriptive Statistics

Table 2 presents the proportions of college graduates who reported a mismatch between his/her job and college major for the full sample and selected subsamples. Among the 5,879 observations, 28.2% reported a major–job mismatch. This proportion is higher than the 20.1% found among US college graduates (Robst, 2007a) and the 20.2% among Swedish college graduates (Nordin et al., 2010).

Robst (2007b, 2008) have discussed the possible reasons for college graduates accepting major mismatched jobs, and the reasons can be placed into two categories: supply–related and demand–related. College graduates may accept mismatched positions for career–related reasons (pay and promotion) or job amenities (working conditions, job location and family related reasons). These reasons can be regarded as supply–related reasons.

Table 1: Variable Definitions

Variable	Definition
Income	Continuous Variable: monthly income from job (including basic wage, bonus, subsidies and all other labor–related income).
Log(Income)	Continuous Variable: natural logarithm of *Income*.
Age	Continuous variable: age at the time of survey.
Tenure	Continuous variable: months in current job.
Mismatch	Unordered categorical variable: 1 if major–job mismatch and 0 otherwise.
Gender	Unordered categorical variable: 1 if male and 0 otherwise.
Homeprov	Unordered categorical variable: 1 if the working place is the respondent's home province and 0 otherwise.
Collprov	Unordered categorical variable: 1 if the working place is in Shandong province and 0 otherwise.
Collcity	Unordered categorical variable: 1 if the working place is the city where the university is located at and 0 otherwise.
Major	Unordered categorical variable: 1 if major category is Natural Sciences, 2 if Engineering, 3 if Agriculture, 4 if Medicine, 5 if Economics, 6 if Management, 7 if Law, 8 if Education, 9 if Literature, 10 if History or Philosophy.
Ownership	Unordered categorical variable: 1 if state–owned enterprizes, 2 if private firms, 3 if foreign companies or joint ventures, 4 if governmental and research institutes and 5 if other job providers.
Provtype	Unordered categorical variable: 1 if the working place is in an under–developed province in inland China, 2 if is in a medium–developed province in inland China, 3 if is in a province in the eastern and coastal areas of medium–developed China and 4 if is in a developed province in the eastern and coastal areas.
Industry	Unordered categorical variable: 1 if manufacturing industry, 2 if construction industry (including real estate industry), 3 if financial service industry, 4 if education industry, 5 if information industry (including information transmission, computer service and software industry), 6 if business service sector (including leasing industry), 7 if logistics and transportation industry (including warehousing and storage industry) and 8 if other industries.
Keyuniv	Ordered categorical variable: 1 if the university is one of the "211–Project Universities" and 0 otherwise.
Firmsize	Ordered categorical variable: 1 if number of employees is less than 100, 2 if between 100 and 500, 3 if between 500 and 2000 and 4 if more than 2000.
Citytype	Ordered categorical variable: 1 if the working place is (or located in) a prefecture level city, 2 if the capital of a province and 3 if the city is under direct supervision of the State Council.

Demand–related reasons for accepting such positions may indicate an unavailability of matched jobs in the degree field or job discrimination that prevents the college graduates from finding a suitable position. Unfortunately, in my data, I do not have the information on why the college graduates accepted major–mismatched jobs. However, I believe that these supply–related and demand–related reasons, which have been found to affect US college graduates' job decisions, have also influenced Chinese college graduates' occupation choices. For example, a Chinese college graduate may take a major-mismatched job because a major–matched job cannot meet his/her pay requirements (supply–related reason). The reason could also be that there is oversupply of college graduates from the same major after the college enrolment expansion and a major–mismatched job is unavailable (demand–related reason).

Table 2: Incidence of Major–job Mismatch

Variables	Categories	Percentage (%)	N
Gender:	Male	27.2	3613
	Female	29.7	2266
Keyuniv:	"211–Project"	26.8	977
	Non "211–Project"	28.4	4902
Major:	Natural Sciences	38.0	547
	Engineering	23.2	2051
	Agriculture	42.2	161
	Medicine	6.60	196
	Economics	30.4	565
	Management	27.4	1349
	Law	56.5	214
	Education	43.1	174
	Literature	24.3	610
	History/Philosophy	50.0	12
Industry:	Manufacturing	24.8	1824
	Construction	23.5	480
	Financial Service	38.4	414
	Education	23.5	468
	Information	21.7	508
	Business Sector	32.6	402
	Logistics/Transportation	35.8	232
	Other industries	32.1	1551
	Full Sample	28.2	5879

Among the male college graduates, 27.2% of them have a mismatch. The proportion is slightly higher for females, being 29.7%. Graduates from ordinary colleges have a slightly higher mismatch rate than "211–Project Universities" graduates. For different major categories, the percentages differ widely. Respondents formerly majoring in Medicine are least often mismatched (6.6%), and this major category is commonly believed to provide highly occupation–specific skills. For those majoring in History and Philosophy, the mismatch proportion is 50%. These skills learnt in the two major categories are more general and easily transferrable to major–mismatched jobs. In light of industry distribution, workers in the information industry are least likely to be mismatched. The mismatch proportion for workers in the financial service industry reaches 38.4%, the highest among all industries.

Table 3 displays the summary statistics for the variables used in this study in terms of groups. For the mismatched group, on average, the monthly income is about 92 *yuan* lower than the matched group, which means that a typical matched employee earns about 5% more than his/her mismatched counterpart. Due to the homogeneity nature of the sample, similar characteristics are found for the two groups. The average age is almost the same for the two groups. The tenure on the current job is only two weeks less for mismatched individuals.[7] For both groups, males account for a larger proportion. The percentage of these graduates who were from "211–Project Universities" is 16.9% for the matched sample, and it is only slightly higher than the mismatched group.

In terms of major categories, Engineering and Management are the two primary sources of graduates, and students who majored in Medicine and History/Philosophy occupy the least proportions. With regard to the ownership of working units, the distributions are quite similar for both groups, with private firms being the largest employers of college graduates. An overwhelming majority of students are observed to work in the more developed eastern and coastal areas of China. About half of these graduates are working in prefecture–level cities in China, and around 10% of them are working in those four municipalities directly under the State Council.[8] For the mismatched group, over one-fourth of these

[7]The shorter tenure on the current job can be the result of ability differences since people with lower ability levels generally need longer time to find a job. However, it can also reflect the fact that it takes a longer time for a college graduate to accept a major–mismatched job. See Section 4 for the discussion on the possible endogeneity problem of *Mismatch*.

[8]The four cities are Beijing, Chongqing, Shanghai and Tianjin.

Table 3: Summary Statistics by Group

		Mismatch	N=1656	*Match*	N=4223
Variables		Mean	SD	Mean	SD
Income		1931.49	1112.96	2023.40	1078.70
Log(Income)		7.445	0.474	7.507	0.444
Age		23.842	1.287	23.919	1.291
Tenure		6.856	2.993	7.498	2.825
Gender		0.593	0.491	0.623	0.485
Homeprov		0.743	0.437	0.726	0.446
Collprov		0.724	0.447	0.729	0.445
Collcity		0.298	0.458	0.306	0.461
Keyuniv		0.158	0.365	0.169	0.375
Major	Natural Sciences	0.126	0.332	0.080	0.272
	Engineering	0.287	0.452	0.373	0.484
	Agriculture	0.041	0.198	0.022	0.147
	Medicine	0.008	0.088	0.043	0.204
	Economics	0.103	0.305	0.093	0.291
	Management	0.223	0.417	0.232	0.422
	Law	0.073	0.260	0.022	0.147
	Education	0.045	0.208	0.023	0.151
	Literature	0.089	0.285	0.111	0.312
	History/Philosophy	0.004	0.060	0.001	0.038
Ownership	State-owned enterprize	0.219	0.413	0.243	0.429
	Private firm	0.418	0.493	0.419	0.493
	Foreign firm, joint venture	0.234	0.423	0.213	0.409
	Government sector	0.116	0.320	0.104	0.305
	Others	0.014	0.117	0.021	0.144
Firmsize	Below 100	0.257	0.437	0.242	0.428
	Between 100 and 500	0.175	0.380	0.215	0.411
	Between 500 and 2000	0.247	0.431	0.246	0.430
	Above 2000	0.321	0.467	0.298	0.457
Provtype	Underdeveloped; Inland	0.010	0.101	0.009	0.096
	Medium–developed; Inland	0.046	0.211	0.054	0.227
	Medium–developed; Coastal	0.016	0.124	0.016	0.125
	Developed; Coastal	0.928	0.259	0.920	0.270
Citytype	Prefecture–level	0.500	0.500	0.488	0.500
	Provincial Capital	0.389	0.488	0.413	0.492
	Four Municipalities	0.111	0.314	0.099	0.299
Industry	Manufacturing	0.273	0.446	0.323	0.468
	Construction	0.068	0.252	0.087	0.282
	Financial Service	0.096	0.295	0.060	0.238
	Education	0.066	0.249	0.085	0.279
	Information	0.066	0.249	0.094	0.292
	Business Sector	0.079	0.270	0.064	0.245
	Logistics/Transportation	0.050	0.218	0.035	0.185
	Other Industries	0.301	0.459	0.249	0.433

students are working in the manufacturing industry, while for the matched group, the proportion reaches 33.2%. The proportions of graduates in the construction, education and information industries are slightly higher for the matched group than those for the mismatched group.

4 The Sorting–by–Ability Problem

The failure to take unobserved ability into consideration can lead to inconsistent estimates. Robst (2007a) acknowledges that the sorting into matched and mismatched jobs could be the result of ability differences among individuals. A college graduate may be forced into accepting a major–mismatched job because the individual is less able and can not compete with his/her peers. Without controlling for the unobserved ability, the variable $Mismatch$ is potentially endogenous.[9] Unfortunately, I do not have a valid proxy for ability. It is also difficult to find a suitable instrumental variable that is highly correlated with the suspected endogenous variable $Mismatch$ but uncorrelated with omitted ability.

However, I do not consider this sorting–by–ability issue as a serious problem. First, the sample I use is the four–year college graduates who had the same amount of schooling, studied in colleges in the same province and graduated in the same year, which is very homogeneous in nature. Second, I know whether they graduated from national key universities ("211–Project Universities") or not. Generally, students admitted to national key universities are believed to have more innate ability. This information helps distinguish those graduates with more ability from those with lower ability levels. Finally, if students with higher ability levels are more likely to find major–matched jobs, then the mismatch proportion should be lower among college graduates from key universities than those from ordinary ones. For the former group, the mismatch proportion is 26.8%, and it is 28.4% for the latter group. The mismatch proportion is indeed higher among graduates from ordinary universities, but the difference is negligible. The mean–comparison test shows that I fail to reject the null hypothesis that the mismatch proportion is the same for graduates

[9] If major–job mismatch has a negative impact on earnings and less able individuals are indeed more likely to sort into mismatched jobs, then the coefficient estimate of mismatch from OLS regression without controlling for ability differences will exaggerate the negative impact of major–job mismatch on earnings.

from national key universities and those from other universities (p–value=0.304). All these evidence offers the reason why unobserved ability does not constitute a serious problem in obtaining reliable estimates of the effects of major–job mismatch on earnings.

5 Estimation Results

5.1 Two Test Results

Before I discuss the estimation results, I perform two tests to justify my approach here. First, I want to know whether the distributions of monthly income between major–job matched individuals and major–job mismatched individuals are significantly different from one another. For this objective, I select the Li et al. (2009) nonparametric test for equality of distributions with mixed categorical and continuous data. I want to test whether the probability density function of income for major–job matched college graduates is significantly different from that for mismatched individuals. My test result shows strong evidence of the differences in the income density functions between the two groups (p–value=0.009).

Second, I want to test whether the variable of focal interest ($Mismatch$) has a deterministic effect on college graduates' monthly income. If the $Mismatch$ variable is found to significantly affect the levels of monthly income, then I can make a strong argument as to why this variable should be included in the income regression. For this purpose, I employ the test by Racine et al. (2006) to test the statistical significance of the $Mismatch$ variable in my nonparametric regression. I find that the major–job mismatch variable has a statistically significant effect on the income levels (p–value=0.000).

The two test results show the evidence of differences in the income distributions between the major–job matched and the mismatched groups, and also indicate the necessity in including the $Mismatch$ variable on the right–hand side of the income regression since it helps explain the differences in the income levels between the two groups.

5.2 Nonparametric Regression Results

5.2.1 Mean Income Effects of Major–Job Mismatch

Table 4 displays the nonparametric coefficient estimates of the *Mismatch* variable. As mentioned previously, I could obtain a unique estimate for each mismatched individual. With so many estimates, I follow the approach by Henderson et al. (2006) to present the results for those mismatched graduates. I present the nonparametric mean estimate, the coefficient estimates corresponding to the 10th, 25th, 50th, 75th and 90th percentiles of the estimate distribution (labeled $Q10$,$Q25$,$Q50$,$Q75$ and $Q90$). The standard errors are obtained via bootstrapping with 300 replications. I also run an OLS regression of $Log(Income)$ on the control variables aforementioned in Section 3.1 and the squared terms of *Age* and *Tenure*. The OLS coefficient estimates of *Mismatch* are also presented for comparison.[10]

I have two major findings in Table 4. First, the penalty of major–job mismatch is much larger in the OLS model than the nonparametric model. My OLS regression results show that compared with major–job matched individuals, major–job mismatched ones suffer from an average of 6% loss in monthly earnings. However, the nonparametric mean estimate shows that the penalty is quite small, only 1.3%. I use the model specification test by Hsiao et al. (2007) to test whether the OLS model is correctly specified. The test rejects the correctness of the OLS model specification at conventional confidence intervals with a p–value of 0.000. I also compare the two models by examining their performances in fitting the sample. The R^2 for the OLS model is 0.3035. However, the R^2 for the nonparametric

[10]The standard errors are much smaller when using the nonparametric method. This related to the property of the nonparametric approach I use in this paper. Conventional kernel methods rely on the assumption that all variables are continuous in nature. When encountering discrete regressors, one option is to split up the data into a number of cells segregated by the combination of categorical variables and nonparametric regression is then implemented within each cell. This sample-splitting approach leads to a loss of precision due to the reduction in sample size. If the number of cells is large, there may not be enough observations in each cell to obtain reliable estimates. By smoothing categorical variables, we have more observations in doing the estimation. However, the merit of this nonparametric approach does not come without a cost. With categorical predictors used in the regression, additional bias has been introduced, although the coefficient estimates are still consistent. Another key property of the nonparametric approach is that by smoothing categorical variables, the mean squared error will be smaller. We know that the mean squared error is the sum of Squared bias and variance. With more bias in the estimates, the variance should decrease sufficiently enough to make the mean squared error smaller. This means that the standard errors will be smaller.

Table 4: Income Effects of Major–job Mismatch

Measure	All	Male	Female
OLS	–0.0590	–0.0588	–0.0605
	(0.0112)	(0.0160)	(0.0163)
NP Mean	–0.0128	–0.0117	–0.0145
	(0.0026)	(0.0029)	(0.0032)
$Q10$	–0.0707	–0.0668	–0.0725
	(0.0034)	(0.0041)	(0.0044)
$Q25$	–0.0369	–0.0355	–0.0391
	(0.0028)	(0.0032)	(0.0032)
$Q50$	–0.0150	–0.0132	–0.0168
	(0.0027)	(0.0030)	(0.0033)
$Q75$	0.0095	0.0118	0.0063
	(0.0029)	(0.0033)	(0.0037)
$Q90$	0.0405	0.0429	0.0358
	(0.0038)	(0.0043)	(0.0053)

Note: Bootstrapped standard errors are reported in parentheses.

model is higher, reaching 0.5645. This difference in sample fitting is somehow expected since nonparametric regression allows for any kind of interaction among all variables. As a result, the nonparametric model could explain more variation in the dependent variable. As the linear model is misspecified and the nonparametric approach involves no functional form and distribution assumptions, I believe that the nonparametric mean estimate is more reliable.

Second, in contrast to the substantial negative impact of major–job mismatch on earnings found among US college graduates in Robst (2007a) and among Swedish college graduates in Nordin et al. (2010), I find much smaller mean effects of major–job mismatch on the monthly income for Chinese college graduates.[11] I also find that the effects do not differ much between the two genders.

With respect to why the average penalty is much smaller for college graduates in China

[11] One related concern is that for a college graduate new to his/her current job, he/she may not report correctly whether the major is mismatched or not due to the unfamiliarity with working conditions and job requirements. As a way to partially resolve the problem, I restrict my attention to the mismatched individuals with at least a three–month current job tenure and reestimate the effects. The mean estimate is –1.24% for the 1395 individuals. Furthermore, I estimate the effects for those mismatched college graduates who stay in the current job for at least half a year and the mean estimate is –1.17% for the remaining 948 individuals. I find little change in the mean impact of major–job mismatch on earnings.

than those in the United States or Sweden, Chinese universities' emphasis on both general and major–specific skills could possibly offer an explanation.[12] China's undergraduate education in four–year colleges is designed to help students gain both general and major–specific skills and ensure the all-round development of students (Zhong, 2007). Generally, when receiving fundamental major–specific undergraduate education in the first two years, subjects like English, Mathematics, Philosophy and Computer Science and many other general–skill–related subjects are also officially required to be taught. Almost all four–year college students are formally required to pass the national College English Test Band 4 (CET4) and obtain the certificate.[13] Taking College English Test Band 6 (CET6) is not compulsory but it is encouraged in most colleges. Tests of computer skills are also required to be undertaken, although depending on the college's specific requirement, the test can be at the provincial level or national level. Many colleges formally refuse to award the bachelor degree to a student who fails to pass either CET4 or the computer skill test. Furthermore, internships and social practices are encouraged by most colleges (Zhong, 2007). All these ensure that college students can gain general skills that are of help in their later career. In the second half of the four–year undergraduate education, students will receive more intensive major and occupation specific training. Thus, through college education, a comprehensive knowledge structure is developed for college students. While Nordin et al. (2010) attribute the large income penalty from major–job mismatch to the highly specialized tertiary education system in Sweden, the emphasis on both general skills and specific skills education by China's four–year colleges ensures that a student can still adapt to his/her occupational positions quickly even if the job is not major–matched. As a result, on average, the major–job mismatch does not result in substantial income penalties.

[12] One other possible reason is that we cannot distinguish between completely matched and partially matched individuals. With partially matched college graduates being considered as being matched, the wage effects of complete major–job mismatch should be smaller.

[13] The College English Test is a national English as a Foreign Language Test for non-English major college students in China. The test has two levels–Band 4 (CET4)and Band 6 (CET6). Band 4 is mandatory for non–English major college students. Passing CET4 is a prerequisite for taking CET6 which is more difficult but optional.

5.2.2 Heterogeneous Income Effects of Major–Job Mismatch

In this section, I look at the heterogeneous effects of a major–job mismatch on earnings. As evidenced from Table 4, the mean effect of –0.0128 conceals the individual–level differences in the impacts of major–job mismatch. Positive effects are observed for the upper quartiles ($Q75$, $Q90$) while effects are negative for lower quartiles ($Q10$, $Q25$). To better describe the distribution of my nonparametric estimates, I plot the density of estimates using a histogram. Figure 1 depicts the density of coefficient estimates for those with a major–job mismatch, which shows that the effects of a major–job mismatch on monthly earnings are positive for some graduates, but negative for others. The impacts range from –24.10% to 30.28%, and they are visually centering and peaking around 0, indicating a modest impact for an overwhelming majority of the individuals. These effects are between –5% and 5% for 1246 of these 1656 mismatched individuals. 7.79%(=129/1656) of college graduates have the effects greater than 5%, and another 16.97%(=281/1656) have negative effects with magnitude greater than 5%. 32.31%(=535/1656) of the college graduates benefit from a major–job mismatch, indicating a mismatch of college graduates does not necessarily mean misallocation of human resources. While it is commonly believed that major-job mismatched individuals will not be rewarded for their major-specific skills, mismatched individuals with supply-related concerns (such as selecting higher-paying and more career-oriented occupations) may have been partially or fully compensated for the earnings loss (Robst, 2007b; Nordin et al., 2010).

5.2.3 Estimates of Major–Job Mismatch for Selected Mismatched Subgroups

Given that I already have a unique estimate for each mismatched individual, I can summarize the results for each subgroup instead of running separate nonparametric regressions (Eren and Henderson, 2008). For every subgroup I care about, I present the mean effects and the effects at the 25th, 50th and 75th percentiles of the estimate distribution for mismatched individuals (labeled $Q25$,$Q50$ and $Q75$) with bootstrapped standard errors.

The results for different college types and major categories are displayed in Table 5. The mean effects of a major–job mismatch on monthly income are both very small for

Table 5: Income Effects of Mismatch for College and Major Groups

Subgroups	NP Mean	$Q25$	$Q50$	$Q75$
College Type:				
"211–*Project*"	–0.0093	–0.0434	–0.0132	0.0208
	(0.0045)	(0.0051)	(0.0044)	(0.0057)
Non "211–*Project*"	–0.0135	–0.0361	–0.0151	0.0080
	(0.0028)	(0.0029)	(0.0029)	(0.0030)
Major:				
Natural Sciences	–0.0125	–0.0356	–0.0117	0.0153
	(0.0050)	(0.0052)	(0.0050)	(0.0055)
Engineering	–0.0110	–0.0327	–0.0136	0.0101
	(0.0033)	(0.0034)	(0.0034)	(0.0038)
Agriculture	–0.0215	–0.0414	–0.0145	–0.0005
	(0.0059)	(0.0064)	(0.0051)	(0.0063)
Medicine	–0.0296	–0.0387	–0.0234	–0.0042
	(0.0098)	(0.0104)	(0.0083)	(0.0099)
Economics	–0.0137	–0.0419	–0.0171	0.0079
	(0.0048)	(0.0050)	(0.0047)	(0.0056)
Management	–0.0122	–0.0356	–0.0140	0.0108
	(0.0035)	(0.0040)	(0.0037)	(0.0042)
Law	–0.0100	–0.0382	–0.0141	0.0109
	(0.0055)	(0.0060)	(0.0053)	(0.0063)
Education	–0.0286	–0.0679	–0.0255	–0.0018
	(0.0062)	(0.0085)	(0.0064)	(0.0067)
Literature	–0.0096	–0.0367	–0.0146	0.0082
	(0.0048)	(0.0052)	(0.0044)	(0.0056)
History/Philosophy	–0.0131	–0.0550	–0.0138	0.0033
	(0.0225)	(0.0272)	(0.0227)	(0.0275)

Note: Bootstrapped standard errors are reported in parentheses.

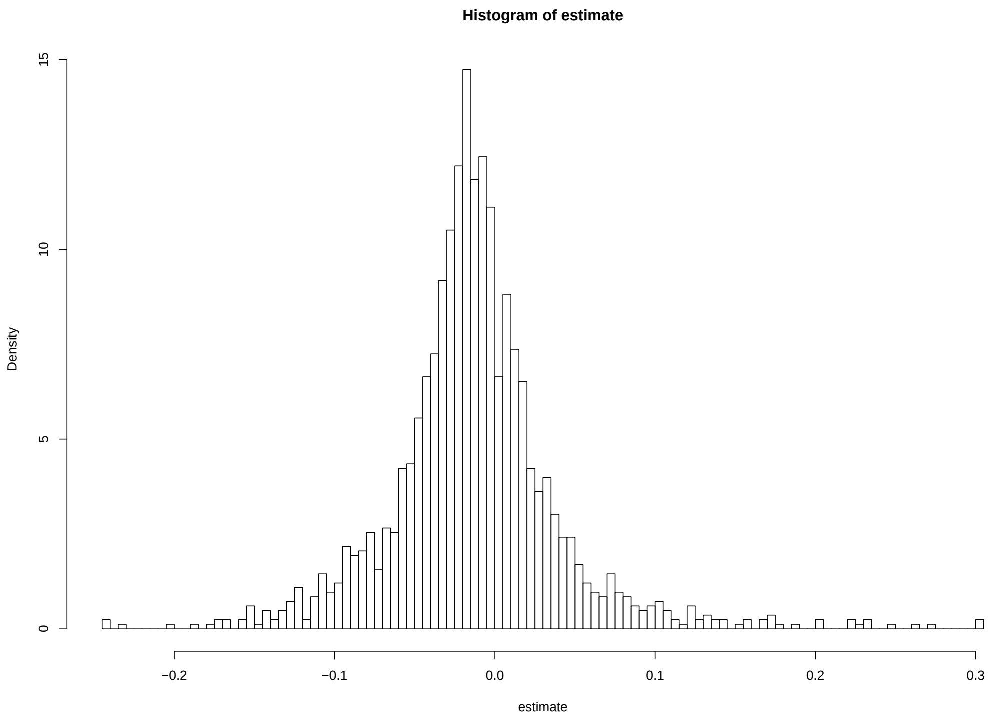

Figure 1: Density of Coefficient Estimates of $Mismatch$ for Mismatched Individuals

national key universities and ordinary universities, although the income penalty is slightly larger for graduates from ordinary colleges. I also find that there is more variation in the effects for key university graduates. For example, the interquartile range ($Q75$–$Q25$) is 0.642 for them, which is larger than the interquartile range of 0.441 for ordinary college graduates.

The mean income effects of major–job mismatch range from –0.0100 to –0.0296 for different major groups. And heterogeneous effects are found within each major group. College graduates from the major category of medicine, , who are commonly known to receive highly occupation–specific skills, have the largest income penalty from major–job mismatch. As the coefficient estimate of $Mismatch$ for medical college graduates is negative at the 75th percentile ($Q75$), an overwhelming majority (over 75%) of college graduates from this major group suffer from income losses. However, even for this group, the average income penalties are much smaller than those reported in previous literature. Among all the major categories, the average income penalty is the minimum for college graduates majoring in literature, which provides relatively more training in general skills.

Estimates for various industries are displayed in Table 6. For workers in the education, information and logistics/transportation industries, the magnitudes of the mean impacts of major–job mismatch are larger than all other industries. College graduates from the education industry are most likely to suffer from the major–job mismatch. Over 75% of workers in the education industry incur the income penalty from a major–job mismatch, although the mean effect is smaller than 3%.

Nordin et al. (2010) have discussed how the phenomenon of education–occupation mismatch is related to the different theoretical approaches that explains how labor markets work, including human capital theory (Becker, 1975) , job competition theory (Thurow, 1975) and the assignment theory (Sattinger, 1993). My findings also provide a test of these existing theories. Human capital theory assumes that a worker is paid by his/her marginal product, and productivity is an increasing function of the worker's human capital investment. This theory implies that workers' earnings are solely determined by worker characteristics from the supply side. As I have found wage differentials among people with the same years of schooling and wage levels are also determined by workers' occupation as well as the match between college major and job, my findings are contradictory to the human capital theory. Job competition theory assumes that it is the characteristics of occupations that determine the productivity and wages. In this model, earnings are related to job characteristics from the demand side and a worker's education does not affect his/her earnings. However, I have found the evidence that one's college education indeed affects the earnings. Thus, my findings do not support the job competition theory either. The assignment theory assumes that the matching of workers and jobs is central. This matching is assumed to influence the productivity and wage of a specific worker–job match. As I find that human capital investment, job characteristics and the match between college major and job are important determinants of wage levels, my findings provide support for the assignment theory by Sattinger (1993).

5.2.4 Coefficient Estimates of Selected Control Variables

Table 7 presents coefficient estimates for selected covariates. *Age* is shown to have little impact on college graduates' mean income determination. The nonparametric mean estimate of *Tenure* indicates that one more month of experience in the current job could increase

Table 6: Income Effects of Mismatch for Industry Groups

Subgroups	NP Mean	$Q25$	$Q50$	$Q75$
Industry:				
Manufacturing	–0.0117	–0.0305	–0.0135	0.0044
	(0.0036)	(0.0058)	(0.0038)	(0.0041)
Construction	–0.0066	–0.0273	–0.0187	0.0213
	(0.0061)	(0.0065)	(0.0058)	(0.0068)
Financial Service	–0.0030	–0.0360	–0.0054	0.0037
	(0.0053)	(0.0060)	(0.0057)	(0.0066)
Education	–0.0255	–0.0468	–0.0246	–0.0085
	(0.0056)	(0.0062)	(0.0057)	(0.0063)
Information	–0.0203	–0.0519	–0.0230	0.0074
	(0.0062)	(0.0070)	(0.0054)	(0.0065)
Business Service	–0.0186	–0.0484	–0.0206	0.0040
	(0.0060)	(0.0059)	(0.0054)	(0.0065)
Logistics/Transportation	–0.0220	–0.0517	–0.0271	0.0053
	(0.0065)	(0.0073)	(0.0065)	(0.074)
Other industries	–0.0127	–0.0320	–0.0135	0.0104
	(0.0039)	(0.0052)	(0.0040)	(0.0044)

Note: Bootstrapped standard errors are reported in parentheses.

monthly income by about 2%. My model also shows that on average males earn 5.25% more than females. In addition, positive coefficient estimates of *Male* are observed at all quartiles, showing being a male pays a premium. The premium to receiving education from a national key university is an 8% increase in monthly income.

I also estimate the impact of the categorical variable *Firmsize* on monthly earnings. My finding that larger establishments pay more is consistent with the literature on the employer size–wage effect (Brown and Medoff, 1989; Gibson and Stillman, 2009). An establishment with more than 2000 workers on average pays about 16% more than the one with 100 workers or less (reference category). Compared with the reference group, an employee earns 4.5% or 8% more if the individual is working in an establishment with 100 to 500 employees or 500 to 2000 workers, respectively. Brown and Medoff (1989) propose several possible reasons for the employer size–wage effect and using US data they favored the explanation that large employers hire higher–quality workers and higher–quality workers earn higher wages. Due to the homogeneous nature of my sample, I find no evidence to support this explanation.

Table 7: Estimates of Selected Control Variables

Variables	OLS	NP Mean	$Q25$	$Q50$	$Q75$
Age	0.0035	–0.0012	–0.0157	0.0003	0.0151
	(0.0041)	(0.0034)	(0.0033)	(0.0034)	(0.0046)
$Tenure$	0.0184	0.0196	0.0119	0.0198	0.0272
	(0.0019)	(0.0015)	(0.0016)	(0.0016)	(0.0017)
$Gender$	0.1140	0.0525	0.0235	0.0542	0.0821
	(0.0116)	(0.0044)	(0.0047)	(0.0048)	(0.0059)
$Keyuniv$	0.1203	0.0813	0.0124	0.0713	0.1355
	(0.0160)	(0.0072)	(0.0062)	(0.0068)	(0.0093)
$Collprov$	–0.2921	–0.2627	–0.3078	–0.2669	–0.2230
	(0.0131)	(0.0089)	(0.0103)	(0.0081)	(0.0091)
$Firmsize$:					
$(0, 100]$	—	—	—	—	—
$(100, 500]$	0.0801	0.0422	0.0042	0.0396	0.0748
	(0.0154)	(0.0053)	(0.0054)	(0.0056)	(0.0064)
$(500, 2000]$	0.1205	0.0765	0.0295	0.0707	0.1180
	(0.0168)	(0.0074)	(0.0067)	(0.0068)	(0.0071)
Above 2000	0.2352	0.1549	0.0867	0.1424	0.2114
	(0.0200)	(0.0096)	(0.0087)	(0.0083)	(0.0108)

Note: Bootstrapped standard errors are reported in parentheses.

6 Concluding Remarks

In this paper, using a recent data set from China, I estimate the effects of major–job mismatch on four-year college graduates' monthly earnings. My nonparametric local linear model shows that, on average, the mismatched individuals suffer from only 1.3% loss in monthly income, a much smaller income penalty than those penalties previously found in developed countries (Robst, 2007a; Nordin et al., 2010). One distinguished feature of the nonparametric method lies in its ability to estimate the effect for every mismatched individual. While the mean impact is negative, 32.31% of them actually benefit from the major–job mismatch. This finding has not been documented before, but it is consistent with the discussion in Robst (2007b). Robst (2007b) has argued that workers accepting a mismatched position due to supply–related factors such as career oriented reasons (pay or promotion opportunities) may earn higher wages than well matched workers.

I further summarize the impact of major–job mismatch on earnings for different subgroups. While the effects vary with the gender, college types, major categories and industries, for almost all subgroups, the mean impacts are negative but less than 3% in magnitude. I find that, compared with other majors, those college graduates who majored in Medicine, which provides highly occupation–specific skills and training to students, are more likely to suffer the income loss from major–job mismatch. The average income penalty is the minimum for college graduates majoring in literature, which provides relatively more generally skills. Those mismatched individuals who work in the Education industry are also more likely to incur an income loss than all other industries. My results also provide a test of the theories that could explain the existence of education–occupation mismatch in labor markets. Among the human capital theory, job competition theory and the assignment theory, my finding that the wage levels are dependent on college education, job characteristics and the match between them provides support for the assignment theory only.

References

Aitchison, J. and C. Aitken (1976). "Multivariate binary discrimination by the kernel method". *Biometrika 63*, 413–420.

Allen, J. and R. van der Velden (2001). "Educational mismatches versus skill mismatches: Effects on wages, job satisfaction, and on-the-job search". *Oxford Economics Papers 53*, 434–452.

Bai, L. (2006). "Graduate unemployment: Dilemmas and challenges in China's move to mass higher education". *The China Quarterly 185*, 128–144.

Becker, S. G. (1975). *"Human Capital"*. National Bureau of Economic Research and Columbia University Press.

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

Buchel, F. and A. Mertens (2004). "Overeducation, undereducation, and the theory of career mobility". *Applied Economics 36*, 803–816.

Chen, D. Y. (2005). "China's mass higher education: Problem, analysis and solutions". *Asia Pacific Education Review 5*, 23–33.

De Grip, A., H. Bosma, D. Willems, and M. van Boxtel (2008). "Job–worker mismatch and cognitive decline". *Oxford Economics Papers 60*, 237–257.

Dolton, J. P. and A. Vignoles (2000). "The incidence and effects of overeducation in the UK graduate labor market". *Economics of Education Review 19*, 179–198.

Eren, O. and J. D. Henderson (2008). "The impact of homework on student achievement". *Econometrics Journal 11*, 326–348.

Freeman, B. R. and A. D. Wise (Eds.) (1982). *"The Youth Labour Market Problem: Its Nature, Causes and Consequences"*. University of Chicago Press.

Gibson, J. and S. Stillman (2009). "Why do big firms pay higher wages? Evidence from an international database". *Review of Economics and Statistics 91*, 213–218.

Hartog, J. (2000). "Overeducation and earnings: Where are we, where should we go?". *Economics of Education Review 19*, 131–147.

Hayfield, T. and J. S. Racine (2008). "Nonparametric econometrics: The np package". *Journal of Statistical Software 27*, 1–32.

Henderson, J. D., A. Olbrecht, and S. W. Polachek (2006). "Do former athletes earn more at work? A nonparametric assessment". *Journal of Human Resources 41*, 558–577.

Hsiao, C., Q. Li, and J. Racine (2007). "A consistent model specification test with mixed categorical and continuous data". *Journal of Econometrics 140*, 802–826.

Korpi, T. and M. Tahlin (2009). "Education mismatch, wages and wage growth: Overeducation in Sweden, 1974-2000". *Labour Economics 16*, 183–193.

Li, Q., E. Maasoumi, and J. Racine (2009). "A nonparametric test for equality of distributions with mixed categorical and continuous data". *Journal of Econometrics, 148*, 186–2000.

Li, Q. and J. Racine (2004). "Cross-validated local linear nonparametric regression". *Statistica Sinica 14*, 485–512.

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

McGoldrick, K. and J. Robst (1996). "Gender differences in overeducation: A test of the theory of differential overqualification". *American Economic Review 86*, 280–284.

McGuinness, S. (2006). "Overeducation in labor market". *Journal of Economic Surveys 20*, 387–418.

Nordin, M., I. Persson, and R. Dan-Olof (2010). "Education–occupation mismatch: Is there an income penalty?". *Economics of Education Review 29*, 1047–1059.

Pagan, A. and A. Ullah (1999). *"Nonparametric Econometrics"*. Cambridge University Press.

Quinn, A., M. and S. Rubb (2005). "The importance of education–occupation matching in migration decisions". *Demography 42*, 153–167.

Racine, J., J. Hart, and Q. Li (2006). "Testing the significance of categorical predictor variables in nonparametric regression models". *Econometric Reviews 25*, 523–544.

Robst, J. (2007a). "Education and job match: The relatedness of college major and work". *Economics of Education Review 26*, 397–407.

Robst, J. (2007b). "Education, college major, and job match: Gender differences in reasons for mismatch". *Education Economics 15*, 159–175.

Robst, J. (2008). "Overeducation and college major: Expanding the definition of mismatch between schooling and jobs". *The Manchester School 76*, 349–368.

Ryan, P. (2001). "The school-to-work transition: A cross–national perspective". *Journal of Economic Literature 39*, 34–92.

Sattinger, M. (1993). "Assignment models of the distribution of earnings". *Journal of Economic Literature 31*, 831–880.

Sicherman, N. (1991). "Overeducation in labor market". *Journal of Labor Economics 9*, 101–122.

Thurow, L. C. (1975). *Generating inequality.* New York: Basic Books.

Wolbers, H. M. (2003). "Job mismatches and their labour–market effects among school-leavers in Europe". *European Sociological Review 19*, 249–266.

Zhao, L. and S. Sheng (2008). "Fast and furious: Problems of China's higher education expansion". East Asia Institute Background Brief No. 395, National University of Singapore.

Zhong, Y. (2007). "Globalization and higher education reform in China". Northeast Normal University, China. Paper source: www.aare.edu.au/05pap/zho05780.pdf.