Are Public Sector Workers Underpaid?

Appropriate Comparators in a Developing Country

Sarah Bales

Martín Rama

This paper was written for the research project on Efficient Public Sector Downsizing (RPO 683-67), with support from the Vietnam Country Office, East Asia and Pacific Region. Advice on econometric issues by Jyotsna Jalan, Michal Lokshin and Quentin Wodon is gratefully acknowledged. Correspondence related to the paper should be sent to mrama@worldbank.org.

Abstract

How to align public sector compensation with the market? In industrial countries, a common reference is the salary paid by private employers for similar jobs. But comparable jobs are formal, whereas in developing countries the relevant alternative for many public sector workers is in the informal sector. Another approach uses as a reference the earnings of similar workers in the private sector, regardless of whether their jobs are formal or informal. A potential shortcoming of this second approach is that workers may differ along characteristics that are unobservable. We assess the importance of this shortcoming by relying on five econometric methods, four of which "correct" the bias from unobservable characteristics. The focus of the paper is on state-owned enterprises in Vietnam. Recruitment into these enterprises was explicitly based on political loyalty and other unobservable characteristics of the workers. A massive downsizing program, whereby the most entrepreneurial workers left, may have exacerbated the selection bias. However, all the results obtained with the "workers approach" fall within a relatively narrow range. They suggest that workers in state-owned enterprises are overpaid by 20 percent or more. The "jobs approach", on the other hand, indicates that they could earn 2 to 6 times more in the private sector.

1. Introduction

Making the public sector more efficient is one of the most formidable challenges faced by policymakers all over the world. In the case of developing countries, the significance of this challenge is revealed by the sectoral distribution of World Bank credits and loans.[1] In the fiscal year 2000, public sector reform was the most important sector of lending, amounting to almost 1.9 billion dollars, out of a total of roughly 15.3 billion. Lending for public sector reform exceeded the combined lending for education and the environment. It also exceeded the combined lending for water supply, sanitation, population, nutrition and health programs. And it was bigger than lending for social protection, which was at one of its highest levels in World Bank history. The only sector that came close to public sector reform was finance. But this is hardly surprising in the aftermath of the East Asian crisis, when many commercial banks still needed to be re-capitalized.

As the wage bill is the largest item in public sector spending, reforming the public sector usually entails changes in employment and pay. This is a socially and politically charged issue. Public sector workers are among the most vocal and influential interest groups in any society. Policy measures that affect them adversely are often met by strong resistance, derailing reform programs and causing governments to fall. Typically, public sector workers claim that they are underpaid compared to their private sector counterparts, argue that low pay is at the root of inefficiency and corruption, and try to re-orient public sector reform in the direction of pay raises. That specific groups of public sector workers (especially at the professional and managerial levels) may be underpaid is not questioned. However, the very fact that most public sector workers are unwilling to leave their jobs, except in exchange for generous compensation, suggests that overpayment is also common.

---

[1] Throughout the paper, the expression "developing countries" is meant to include transition economies as well. These countries and economies may differ in their output per capita, but one of the key features they have in common is the sizeable portion of the labor force that is self-employed or not subject to labor market regulations.

The absence of reliable information on the gap between public and private sector pay may increase the leverage of interest groups to influence government decisions regarding public sector pay. This lack of information is also a hindrance when designing downsizing programs. Compensation packages for redundant workers tend to be set up in ad hoc ways. Typically, some rule of thumb involving salary and seniority in the public sector is used, but the resulting amount of compensation bears no relationship with the present value of the loss in earnings and benefits from job separation. While many separated workers have been under-compensated, some of the separation packages used in developing countries could easily qualify as "golden handshakes" (Kikeri, 1997). Downsizing programs supported by the World Bank are not an exception in this respect (Haltiwanger and Singh, 1999).

Unfortunately, the method that is most frequently used to estimate the gap between public and private sector pay is bound to produce biased results in developing countries. This method, called "the jobs approach" in what follows, focuses on the salaries of a set of private sector jobs whose description is similar to that of public sector jobs. A usual version of the jobs approach is the "Hays points" system used by many large organizations (including, until quite recently, the World Bank itself) to set their compensation levels. The problem when applying this approach to developing countries is that comparable jobs are mainly or exclusively found in formal sector enterprises. Jobs of this sort might be the relevant alternative for public sector workers at the professional and managerial levels. However, tracer studies of separated public sector workers suggest that the relevant alternative for the rank-and-file is self-employment, or casual work in informal activities, including agriculture (Alderman *et al*., 1996, Rama and MacIsaac, 1999). As salaried jobs in formal sector enterprises are among the best in a developing country, the jobs approach overestimates the extent to which public sector workers are underpaid, or underestimates the extent to which they are overpaid. Despite its obvious bias, this approach is often used in World Bank-supported reform programs.

The alternative is to compare the earnings of public sector workers to those of similar workers in the private sector, regardless of whether they are employed in the formal or the informal sector of the economy. This method is called "the workers approach" in what follows. One potential problem with the workers approach is that

individuals may be similar along some observable dimensions (e.g., gender, age or educational attainment) but different in more subtle ways. For instance, public sector workers could be more talented, or less ambitious, or better connected, than their private sector counterparts. Comparing the earnings of a public sector worker to those of an apparently similar private sector worker may thus be misleading. What needs to be evaluated is the earnings a public sector worker would have if he or she were to move to the private sector, taking into account his or her talent, ambition or connections. But those earnings are unobservable. By ignoring relevant unobservable characteristics, the workers approach could therefore lead to results that are as biased as those obtained with the jobs approach, except that the direction of the bias is unknown a priori.

The goal of this paper is to assess whether the bias created by unobservable individual characteristics, also known as selection bias, is substantial enough to invalidate the workers approach. For this assessment to be credible, it has to focus on a country that satisfies two criteria. First, its database has to be good enough to credibly measure the gap in earnings and benefits between the public and the private sector, "corrected" for the effect of unobservable individual characteristics. The availability of a nationally representative household survey with a panel structure over time is the ideal in this respect. Second, there has to be a strong presumption that recruitment into the public sector, and separation from it, is influenced by unobservable individual characteristics that influence private sector earnings. If, say, recruitment were strictly based on educational attainment, it would not be surprising to find that the selection bias is negligible. But this finding would be irrelevant for other, less meritocratic countries.

Vietnam meets these two criteria. As regards the public sector, it was government policy to use jobs in state-owned enterprises (SOEs) as an income transfer. In principle, some of these jobs were to be considered as a reward for political loyalty or a compensation for sacrifices incurred during independence wars. In practice, many of them were allocated based on connections. Presumably, those who got the jobs would have had lower earnings in the private sector. But political loyalty, war-related losses or connections are not measurable. More recently, SOEs were subject to a massive downsizing program that led to the separation of roughly one third of their workforce in the early 1990s, partly on a voluntary basis. Those who remained were probably less

entrepreneurial, and had worse earnings opportunities in the private sector. Again, the entrepreneurial spirit is not measurable. Because of the role played by unobservable characteristics in both recruitment and separations, the selection bias could be considerable when estimating the potential earnings of SOE workers in the private sector.

Concerning the data, Vietnam has a high-quality, nationally representative survey with a panel structure, known as the VLSS (for Vietnam Living Standards Survey). The VLSS reports detailed information on sector of employment and earnings, including a vast array of benefits and payments in kind, for a large sample of workers. It was carried out in 1992-1993 and in 1997-1998, and a substantial fraction of the workers was interviewed in both rounds. Because the public sector of Vietnam stills employ a non-negligible portion of the labor force, in each of the two rounds of the VLSS it is possible to find hundreds of workers whose main occupation is in an SOE. The panel nature of the VLSS also makes it possible to "remove" the effect of unobservable characteristics, by focusing on the change in earnings experienced by individuals who moved from SOEs to the private sector, or vice-versa. But panel data analysis is only one among several econometric techniques that can be applied to VLSS data to correct for the effects of selection bias.

Our paper is certainly not the first one to use the workers approach, or to explicitly assess the impact of selection bias on the earnings gap between the public and the private sectors in a developing country. To our knowledge, studies along similar lines exist for Côte d'Ivoire (van der Gaag and Vijverberg, 1988), Ethiopia (Mengistae, 1998), Haiti (Terrell, 1993), India (Lakshmanasamy and Ramasamy, 1999), Indonesia (Filmer and Lindauer, 2001), Peru (Stelcner *et al.*, 1989), Poland (Adamchik and Beri, 2000), Taiwan (Hou, 1993) and Tanzania (Lindauer and Sabot, 1983). Another study uses the workers approach to estimate the earnings gap for a specific group of public sector workers, namely teachers, across a dozen Latin American countries (Psacharopoulos et al., 1996).

However, our paper differs from previous studies in several, important ways. First, it focuses on the entire distribution of the earnings gap, and not only on its "average" size. Second, rather than choosing one method to correct for the selection bias,

it relies on a broad array of empirical strategies. A valid criticism of all the methods used to correct for the effect of unobservable individual characteristics is that they require strong assumptions. By comparing the results obtained with alternative methods, hence under alternative assumptions, this paper implicitly evaluates the robustness of the estimated distribution of earnings gaps. Third, the paper also compares the results obtained with the workers approach to those obtained with the jobs approach. This comparison reveals how misleading the latter can be.

2. State-Owned Enterprises in Vietnam

Although the public sector of Vietnam is not large compared to other transition countries, it still plays a significant role in the economy. SOEs employ roughly five percent of the labor force, but account for around a fifth of GDP. This higher productivity is the result of a much higher capital intensity of production, which in turn may require the use of relatively more qualified workers. Technology is only one among many reasons why SOE workers differ from private sector workers. SOE workers are older on average than private sector workers, as the public sector expanded in the times of central planning, but has been shrinking since market-oriented reforms (or *Doi Moi*) began, in the late 1980s. As Vietnam is a young country, most of the new entrants to the labor force end up in the private sector. In a country that is still massively rural, SOE workers are also more urban. And SOE workers are predominantly male, despite the fact that labor force participation rates are similar for men and women in Vietnam. This male bias exists not just compared to the private sector in general, but also compared to formal enterprises in the private sector (MOLISA, 1998).

Public sector workers differ from private sector workers in more subtle ways too. In Vietnam, it has been an explicit government policy to use SOE jobs as an income transfer. Shortly after the French war, it was decided that "people to be recruited must be from all sectors: northern, southern, female, male, with priority on recruiting people who had been active or had achievements in the war and are working to build peace" (circular 8/LD-TT of August 1959; our translation). While skill levels were deemed important, the required standards included "political quality, workers disposition [and] health", all characteristics that might be difficult to measure using a household survey instrument.

Political quality might have been especially relevant, as all large SOEs have a Communist Party cell. One regulation explicitly stated that "recruitment for technical training must be combined with arranging work for children and other relatives of existing cadres, of soldiers, of war martyrs, of revolutionaries" (circular 2-LD/TT of January 1960; our translation). This regulation also criticized the common practice of "hiring friends and family without the necessary skills" (ibid.), which suggests that connections were an important determinant of recruitment. Additional regulations giving priority in recruitment to demobilized soldiers, war invalids and their families were issued during the American war.

Another potential source of selection bias was the massive downsizing program of the early 1990s, when roughly one in three SOE workers left the state sector. By then, the *Doi Moi* program had created a significant earnings potential in the informal sector. The severance pay package offered to those resigning was not too large, but it was considered sufficient to start a small household business. It would not be surprising if many among the most entrepreneurial SOE workers had left at that time. The downsizing regulations did not specify that separations had to be voluntary, but they gave considerable leeway for SOEs to manage the process. Decision 176-HDBT, of October 1989, explicitly encouraged the separation of "young workers who are healthy and have skills, if they volunteer" (our translation). Admittedly, the share of female workers who lost their jobs was disproportionately high, and this could be an indication that many separations were involuntary (Rama, 2001). However, the downsizing program of the early 1990s would be a potential source of selection bias as long as some of the separations were voluntary, and many of them were.

Despite this massive downsizing, SOEs remain substantially over-staffed in Vietnam. An analysis based on plant-level data suggests that as many as half of the workers would be redundant if SOEs were to operate in the same way as fully private enterprises (Belser and Rama, 2001). Redundancies are especially large in construction, mining and transportation. They are smaller, or even negligible, in footwear, textile and garments. When this paper was written, the government of Vietnam was initiating a new phase of its reform program, geared towards the liquidation, privatization or restructuring of several thousand SOEs. This new phase of the program could require one hundred

thousand job separations per year, in addition to the "natural" attrition resulting from retirement and contract expiration. Predicting the losses that SOE workers could experience as a result of job separation is key to designing an appropriate compensation and assistance scheme for them.

3. The Jobs Approach

A common perception, after almost a decade of reforms, is that SOE workers are underpaid compared to their private sector counterparts. According to a survey carried out by the Vietnam General Confederation of Labor, "state economic units in Hanoi have income on average of 450,000 dong per person per month; non-state economic entities: 500,000 dong per person per month; foreign invested enterprises: 90 US dollars [equivalent to 993,000 dong] per person per month" (*Labor*; October 15, 1996, page 3; our translation). Another study, by the Ministry of Education and Training concluded that "self-employed workers have incomes 26 percent higher than state workers, people who work in private companies have incomes 33 percent higher, and people working in foreign companies have incomes 73 percent higher (*Labor*, 21 December, 1996, page 1; our translation). Not surprisingly, university graduates are said to be lured into the private sector to the detriment of SOEs. According to one report, "starting salaries of between 200 and 300 US dollars per month were on offer compared to only 20-50 dollars a month in state-firms" *(Courier du Vietnam*; May 30, 1997; our translation).

This perception is confirmed when the jobs approach is used to estimate the gap in earnings between the public and the private sector in Vietnam, as shown by Table 1. The first two data columns in this table report the average earnings, including benefits, of workers in selected occupations according to individual records from the 1997-98 round of the VLSS. The workers in the first data column are employed by SOEs; those in the second data column have jobs in the private formal sector. For the purpose of this table, a private sector job is defined as formal if the firm has a personnel of ten or more and the worker has a written contract.

The last two data columns in Table 1 report earnings for similar jobs with the "best" employers in Vietnam, according to a study commissioned by the United Nations

Development Program (UNDP). The employers considered include local subsidiaries of firms such as Citibank and British Petroleum, among others. This study was aimed at determining the appropriate pay level for local UNDP staff, not for public sector workers. Still, the kind of pay study that would be carried out in the context of a public sector reform program would probably be quite similar to the one commissioned by the UNDP.

The matching of occupations across the VLSS survey and the UNDP study is probably not perfect. Several jobs had to be discarded, either because they were not similar enough or because they included very few observations. Still, the picture that emerges from Table 1 would not be substantially different under different matching hypotheses.

According to Table 1, SOE workers are clearly underpaid. If the figures in this table are to be taken literally, an office worker would earn more than twice as much in the formal private sector, and five to six times as much with the "best" employers in Vietnam. The gaps vary from occupation to occupation, but the basic conclusion holds. If SOE wages were to be revised based on the jobs approach, the obvious recommendation would be to raise them substantially. However, the difference in the size of the recommended wage increase, depending on which comparator is used, casts doubts on the reliability of this approach. If resources were not a constraint, should the wages of office workers in SOEs be multiplied by a factor of two, or by a factor of six? In fact, the workers approach suggests that both figures are wrong. But its implementation requires some explanation.

4. The Workers Approach

As a first approximation, the gap between the salary in the SOE and the alternative labor earnings in the private sector can be estimated using a very simple econometric model. Let X_i be a vector representing the observable characteristics of worker 'i' (e.g., gender, educational attainment, work experience, etc.) and W_i be the labor earnings of this worker. The sector of employment can be captured by the indicator variable S_i, which is equal to one if worker 'i' is employed by an SOE and equal to zero if he or she is in the private sector. The following relationship between individual

characteristics and labor earnings can be assumed for those who work in the private sector:

$$Log\,W_i = a_X\ X_i + e_i \text{ for } S_i = 0 \qquad (1)$$

where e_i is a stochastic disturbance with zero mean. This disturbance summarizes the effects of unobservable individual characteristics, such as talent, political loyalty, war-related losses, or connections. Assuming that e_i is not correlated with the unobservable individual characteristics X_i, the parameters in vector a_X can be estimated by ordinary least squares.

The estimated parameters, identified in what follows by a hat, can in turn be used to predict the earnings an SOE worker with individual characteristics X_i would have in the private sector. More specifically, the gap between actual earnings in the SOE sector and predicted earnings in the private sector can be defined as:

$$R1_i = Log\,W_i - \hat{a}_X\ X_i \text{ for } S_i = 1 \qquad (2)$$

The gap in equation (2) is identified by the number one to indicate that it is estimated with the first (and simplest) econometric model considered in this paper. For relatively small values of this gap, R1 can be interpreted as a percentage.

The model just described generates biased results when labor earnings are affected by unobservable individual characteristics that are correlated with the sector of employment. The switching regression model is one among several empirical strategies to deal with this problem. This model assumes that the sector of employment is determined as follows:

$$S_i^* = m_X\ X_i + m_Z\ Z_i + h_i \text{ for all i} \qquad (3)$$

$$S_i = 1 \quad if \quad S_i^* > 0$$

$$S_i = 0 \text{ otherwise}$$

In equation (3), Z_i are individual characteristics that affect the latent variable S^*_i, hence potentially the sector of employment, but not the level of earnings.

The model is completed by two earnings equations, one for each sector:

$$Log\,W_i = g_X^0\ X_i + w_i^0 \text{ for } S_i = 0 \tag{4A}$$

$$Log\,W_i = g_X^S\ X_i + w_i^S \text{ for } S_i = 1 \tag{4B}$$

The disturbances ? ^0_i and ? ^S_i are potentially correlated with the disturbance ?$_i$ of the underlying model for sector selection. In fact, the selection bias resulting from unobservable individual characteristics is captured in this model through the correlation between disturbances. For example, politically oriented individuals could be more likely to work in the public sector, and to earn higher salaries in it, but be less productive in a private sector job. In this example, the correlation between ?$_i$ and ? ^S_i would be positive, whereas the correlation between ?$_i$ and ? ^0_i would be negative.

Assuming a joint normal distribution for all the disturbances, the system represented by equations (3), (4A) and (4B) can be jointly estimated using the full-information maximum likelihood method. This procedure yields unbiased values for coefficients μ_X, μ_Z, ?$_X$ and ?$_Z$, as well as for the standard deviations s_0 and s_S of disturbances ? ^{0} and ? ^{S}. It also yields estimates for the correlation coefficients between ? and ? ^{0}, and between ? and ? ^{S}, called ?^{0} and ?^{S}.

The estimated parameters can be used to predict the earnings public sector workers would have in the private sector, taking into account both their observable and unobservable characteristics. The prediction involves the density function $f(.)$ and the accumulated density function $\Phi(.)$ of a normal distribution, evaluated at $\hat{m}_X\ X_i + \hat{m}_Z\ Z_i$. Let f_i and Φ_i be those values. The predicted gap in earnings can be written as:

$$R2_i = Log\,W_i - \hat{g}_X^0\ X_i + \hat{r}_0\ \hat{s}_0 \left(\frac{f_i}{1-\Phi_i} \right) \text{ for } S_i = 1 \tag{5}$$

The interpretation of equation (5) is not straightforward. Consider again the case where the relevant unobservable characteristic is political motivation, and the correlation

coefficient $\hat{r}^0$ is negative. Consider also two SOE workers: one who was recruited mainly because he or she had the right skills (education, experience, etc.) and one that was recruited based on political considerations. The first, skilled worker, is characterized by a high value of the latent variable S* and a high value of the ratio $f/(1-\Phi)$. For the second, politically motivated worker, the ratio $f/(1-\Phi)$ is small. Because $\hat{r}^0$ is negative (and $\hat{s}^0$ is positive), the last term in equation (5) is larger for the politically motivated worker than for the skilled worker. Put differently, based on their unobservable characteristics only, the politically motivated worker loses more if he or she has to move to the private sector.

The next two models considered in this paper involve the use of panel data. Their key assumption is that the effect of the relevant unobservable characteristics on earnings can be found at least twice in the data. In the simplest case, the same individual is observed at two points in time. The labor earnings of this individual can be seen as the outcome of three different determinants: his or her observable individual characteristics, the sector he or she works in, and his or her unobservable characteristics. The effect of the latter, hereafter called n_i, is supposed to be invariant over time and across sectors. The resulting earnings equation is:

$$Log\,W_{it} = d_X\ X_{it} + d_S\ S_{it} + n_i + x_{it} \text{ for all I} \tag{6}$$

where the sub-index t indicates a point in time. Assume, for instance, that a more talented individual earns more in both sectors. In that case, n_i would be positive.

The availability of at least two observations for individual i makes it possible to estimate all the coefficients in equation (6), including n_i using panel data techniques, such as fixed effects or random effects. Having estimated the impact of unobservable characteristics on earnings, the gap between the SOE salary and private sector earnings becomes:

$$R3_{it} = Log\,W_{it} - \hat{d}_X\ X_{it} - \hat{n}_i \text{ for } S_i = 1 \tag{7}$$

In the example, a talented private sector worker would be characterized by a positive value of $\hat{n}_i$, so that private sector earnings would be larger, and the earnings gap smaller, than suggested by observable characteristics only.

The individual effects model is well suited to the case where the relevant unobservable characteristics are indeed individual-specific. However, the sector of employment may also depend on unobservable characteristics that are better seen as household-specific, such as political loyalty, war-related losses or connections. In this case, information on the earnings of two members of the same household can be used to estimate the effect n_h of unobservable characteristics. This model, labeled household effects in what follows, leads to the following earnings equation:

$$Log\,W_{ih} = j_X X_{ih} + j_S S_{ih} + n_h + x_{ih} \text{ for all i and h} \tag{8}$$

where the sub-index h identifies the household. For example, all working members of a well-connected household could have higher earnings than suggested by their observable characteristics only, both in the public sector and out of it. In that case, n_h would be positive, reflecting the premium to connections.

As before, all the coefficients in this equation, including n_h, can be estimated using panel data techniques, such as random effects or fixed effects. Those coefficients can in turn be used to predict the earnings gap as follows:

$$R4_{ih} = Log\,W_{ih} - \hat{j}_X X_{ih} - \hat{n}_h \text{ for } S_{ih} = 1 \tag{9}$$

In the example, a member of a well-connected household who works in an SOE would have a lower earnings gap than suggested by his or her individual characteristics only, because connections would allow him or her to get a good job out of the public sector.

The last model considered in this paper is based on a direct "matching" of public and private sector workers. Unlike the previous models, this one does not attempt to unveil the structure of private sector earnings. The comparison is directly between each

public sector worker and one or several "similar" workers in the private sector, not between a public sector worker and a point in a regression line. Predicting alternative earnings based on a regression line is the same as setting the stochastic disturbance of the earnings equation equal to zero. But this assumption is not necessary with the matching model. The best match for a public sector worker is a worker who, in general, has earnings either above or below the regression line.

The only "structure" used by the matching model can be found in the mechanism that determines the sector of employment, which is still supposed to be the one described in equation (3). This equation is estimated using a Logit model, and the resulting coefficients are used to measure the "distance" D_{jk} between an individual j who works for an SOE and an individual k who works in the private sector. This distance is defined as the square of the difference between the predicted probabilities that these two individuals would work for an SOE:

$$D_{ik} = \left[\frac{\exp(\hat{m}_X X_j + \hat{m}_Z Z_j)}{1+\exp(\hat{m}_X X_j + \hat{m}_Z Z_j)} - \frac{\exp(\hat{m}_X X_k + \hat{m}_Z Z_k)}{1+\exp(\hat{m}_X X_k + \hat{m}_Z Z_k)} \right]^2 \tag{10}$$

Among all the k individuals who work in the private sector, individual n is considered the best match for public sector worker i if $D_{in} < D_{jk}$ for all $k \neq n$. In addition, it is required that the predicted probability of working in an SOE be non-negligible, for both individual i and individual k, and that the distance D_{in} does not exceed some critical threshold. With the matching approach, the gap in earnings is defined as:

$$R5_i = Log\,W_i - Log\,W_n \text{ for } S_i = 1 \tag{11}$$

This model can be easily extended so as to consider more than just one, best match for each SOE worker. In this paper, the average earnings of the best three matches for each SOE worker will be used as the appropriate W_n.

The five models outlined in this section rely on different assumptions to estimate the earnings gap between the public and the private sector. Moreover, as will be discussed below, the sample of observations that can be used in each case is generally different too. The five R indicators are thus the joint outcome of both different

assumptions and different data sets. For instance, the ordinary least squares model does not attempt to correct for the effects of the selection bias, but it uses the largest number of observations. On the other hand, the individual effects model provides a highly credible way to "remove" the effect of unobservable characteristics, but it can only be applied to the fewer individuals whose earnings are observable at two points in time. Rather than trying to argue in favor of one or the other of the R indicators, this paper compares the results obtained with all five. This comparison is key to assess how sensitive the predicted earnings gap is to the chosen econometric technique.

5. Data

The data used in this paper are from the Vietnam Living Standards Surveys (VLSS), collected by the General Statistical Office in Vietnam. The VLSS is a multi-purpose instrument, in the spirit of the Living Standards Measurement Surveys (LSMS) set up in various developing countries with support from the World Bank (Glewwe and Grosh, 1998). While one of the main objectives of these surveys is to estimate household consumption, hence to measure poverty, the VLSS has an especially rich questionnaire regarding individual employment and earnings. In addition, and unlike most surveys for developing countries, the VLSS has a panel component, whereby information on a large number of individuals is available at two points in time. The first VLSS survey was implemented from October 1992 to October 1993, and the second one from December 1997 to December 1998.

For the first VLSS a sample of 4,800 households was selected using multi-stage cluster sampling, stratified by urban and rural areas according to the 1989 population census. The first stage of the sampling systematically selected 120 communes and 30 urban wards, out of all the communes and wards in the country, using a fixed household interval. The second step randomly selected two villages within each of these communes or wards, again in proportion to the number of households. Within each village, a household list was prepared by commune authorities based on officially registered households and 16 households were chosen, with an extra four as reserve in case a household was unavailable or unwilling to be interviewed. The second VLSS had a sample of 6,000 households, of which 4,305 were in the original 1992-93 sample. The

remainder were chosen from the Multi-purpose Household Survey (MPHS) and from replacement households selected during the fieldwork. The second VLSS used a different system of stratification based on three sizes of urban areas and the rural areas of the seven regions that existed at the time of sample selection. Instead of choosing the sample to be self-weighting, as in the first round, smaller regions were over-sampled to guarantee a sufficient number of observations for analysis at the regional level (see Bales, 2000). As a result, data from the second VLSS must be weighted to avoid sampling bias.

The contents of the VLSS household and commune surveys are quite comprehensive, with over 100 pages of questions organized in 15 different sections. The individual data used in this paper are from the sections of the household questionnaire dealing with basic demographic characteristics, education, and labor and employment. Household data from the sections on expenditures and assets are used as controls. Household data are also used to assess the robustness of results, by evaluating the relationship between sector of employment and consumption per capita. Other control variables, such as road and waterway access, and existence of factory and cottage industries at the local level, are from the community questionnaire. Finally, all analyses control for geophysical region, as delta areas have certain advantages over other regions in Vietnam, such as greater ease of transport.

Employment status and labor earnings are the two key variables for the analysis in this paper. The VLSS collects information on primary and secondary employment for all the working members of the household who are six years of age or older. This information refers to occupation, sector of employment, hours worked and various categories of compensation. Unfortunately, data on labor earnings are only available for wageworkers, and not for the self-employed. This leads to an over-representation of the formal sector (the salaried relationship is less prevalent in the informal sector). In this paper, total earnings are calculated on an annual basis, using information on actual hours of work per week and the actual weeks of work per year. Figures are annualized for those who had worked for less than one year at the time of the survey. All earnings figures are adjusted using price indexes that are specific to each region and month of the year.

Three main earnings indicators are considered in the analysis. The first one is the annual compensation received from the primary occupation, including salaries, wages, bonuses, various supplements and payments in-kind. A second indicator is total annual earnings, including payments in cash and in kind, from all forms of wage employment, including primary and secondary occupation as well as other jobs held in the past 12 months. The third one is the average hourly compensation from primary occupation, which is calculated dividing the first earnings indicator by the number of hours worked in the primary occupation over the year. Finally, the section dealing with the robustness of the results also considers a fourth indicator, which is annual consumption per capita.

The other key variable is the sector of employment. The analysis focuses on the differences between the SOEs and the private sector, excluding government agencies. Turnover among civil servants is low in recent years, and no substantial change in their number is foreseen in the near future. Therefore, the relevant alternative for an SOE worker is not a job as a civil servant, but rather a job in the private sector. In the empirical analysis, anyone who reported working in an SOE as his or her main employment in the past 12 months was considered an SOE worker. Those who reported working in a household enterprise, in a collective or cooperative enterprise, in a private firm or a foreign-owned company, as well as those who reported doing casual work, were considered private sector workers. Those who reported working for the government (including teachers, doctors and administrators), or for a mass organization, were excluded from the sample.

Table 2 presents summary statistics for the variables used in the econometric analysis, disaggregated by sector of employment. The figures correspond to the 1997-98 VLSS, as this paper estimates the earnings gap for the most recent period only. The top panel of table 2 describes the four earnings indicators considered in the paper, including consumption per capita. These are the W variables in the analysis. It appears that SOE workers earn almost twice as much as private sector workers, but this crude comparison does not take into account that they also differ in other ways. The next panel report on the observable characteristics of individuals and the households they belong to. Household characteristics are used as controls when the earnings indicator is consumption per capita.

The last panel contains information on the characteristics of the communities individuals live in. The bottom three panels of table 2 correspond to the X variables in the analysis.

6. Model Estimation

Four of the five econometric methods used in this paper involve the estimation of an earnings function for private sector workers. The results are presented in Tables 3 to 6. In terms of the methodology section above, the reported coefficients are the estimated values of parameters a_X (in Table 3), g_X^0 (in Table 4), d_S and d_X (in Table 5) and j_S and j_X (in Table 6). The three columns in each of these tables correspond to each of the three earnings indicators used as dependent variable in the analysis. The explanatory variables include observable individual characteristics and community characteristics. The fit of all these regressions is satisfactory by conventional standards. The sign and magnitude of the estimated coefficients is consistent with results obtained in other countries. If anything, the coefficient on the number of years of education is small by international standards, but this is hardly surprising in a country in transition from central planning to a market economy.

Some of the econometric methods used in this paper involve the estimation of other relationships, not reported in Tables 3 to 6. In particular, the switching regression model requires the simultaneous estimation of another earnings function, for SOE workers, and a participation model, explaining who is employed by an SOE and who works in the private sector. The explanatory variables of the underlying participation model include all the individual and community characteristics affecting private sector earnings. But they also include a set of additional variables, which was identified as Z_i in the methodology section of the paper. Among these additional variables are household characteristics (the same ones that were listed in the third panel of Table 2) and the share of SOEs in total employment at the province level. This share was estimated based on the 1997-98 round of the VLSS, taking into account all workers, regardless of whether they were wage earners or not.

The additional explanatory variables used to estimate the participation model are supposed to affect the sector of employment, but not the level of earnings within each

sector. For instance, individuals should be more likely to work in the public sector in communities where the latter is bigger. But pay scales in the public sector should not be affected by the share of SOEs in total employment. And pay in the private sector should not be affected either, as it is largely determined by labor productivity. The same set of additional variables was considered when estimating the participation equation for the matching model.

Results based on panel data techniques involve a smaller number of observations and, in some cases, a reduced set of explanatory variables. In the individual effects model, the number of observations shrinks because data on earnings in 1992-93 are not available for all the individuals who had earnings in 1997-98. Comparisons over time can be carried out only for those individuals whose earnings are observed twice. In the household effects model, there are fewer observations because many households have only one wage earner. Comparisons within households cannot be implemented in their case.

Several reasons explain the decline in the number of explanatory variables. In the case of the individual fixed effects model, the number of years of vocational training received had to be dropped, because it was measured differently in the two rounds of the VLSS. In the case of the households effects model, the restriction resulted from the econometric technique that had to be applied. When using panel data, the choice of the appropriate technique is determined by the Hausman test. The most efficient technique is random effects, but it produces biased results when the explanatory variables are correlated with the disturbance of the equation. The Hausman test indicates they are indeed correlated in the case of the household effects model, thus requiring the use of fixed effects. With this technique, all the variables that take the same value for all individuals in a household (i.e., all community characteristics and regional dummies) need to be dropped. Hence the much smaller number of explanatory variables.

For the matching model, a first step was to exclude from the sample those private sector workers who appeared to be “too” different from SOE workers. This was achieved by predicting the probability of being employed by an SOE for all workers, regardless of their sector of employment. Two distribution functions were then drawn for this predicted

probability: one for workers whose actual job was in an SOE, and another one for workers whose actual job was in the private sector. Only the predicted probabilities which had some density in both distributions (also called the "common support") were retained. In practice, this amounted to discarding less than five percent of private sector workers.

Also, in order go ensure greater uniformity an effort was made to match workers within each geographical area. As some of the areas contained too few observations, it was necessary to combine the northern mountains and north-central coast regions, as well as the South Central Coast, Central Highlands and Mekong Delta regions. However, the Red River Delta and Southeast were left separate. For each of these regional groups, matching was done separately for rural areas, small and medium areas, and the two biggest cities. Pairs of observations separated by a distance of 0.02 or more were dropped. Likewise, any case which had less than 3 matches was also dropped.

7. Predicted Earnings Gaps

The distribution of the individual earnings gaps estimated with each of the five models is summarized in Tables 7 to 9. Each of these tables corresponds to one of the earnings indicators considered. The bottom portion of the tables reports basic statistics on the estimated earnings gaps, considered one at a time. The top portion reports correlation coefficients between pairs of estimated earnings gaps. Several conclusions emerge from these tables.

First, it appears that workers whose main occupation is in an SOE have substantially higher annual earnings in their main job, and in all jobs, than if they had to move to the private sector. According to tables 7 and 8, the mean earnings gap is estimated at around 20 percent by all but one of the models that do correct for selection bias. The exception is the individual effects model, whose mean earnings gap appears to be twice as high. The mean earnings gap estimated without correcting for selection bias falls in between. A similar pattern is observed regarding the median earnings gap. It follows that a majority of SOE workers earn more than their private sector counterparts,

given observable and unobservable characteristics. This first conclusion is at odds with the results obtained when using the jobs approach.

The gap in earnings in favor of SOE workers is not due to a substantially higher remuneration per hour of work, however. Table 9 suggests that hourly earnings for the median SOE worker are within 5 percent of what he or she could make in the private sector. Again, estimates based on the individual effects model are the exception, as they indicate that hourly earnings are about 20 percent higher in SOEs than in the private sector. And estimates without correcting for the effect of selection bias fall in between. But overall, based on Table 9 it is difficult to claim that hourly earnings in SOEs are dramatically out of line with the private sector. The main difference is that SOE workers get remunerated for a substantially higher number of hours, compared to their private sector counterparts.

A third conclusion is that the earnings gaps estimated using the jobs approach is not supported by any of the econometric models used in this paper. Based on Table 1, private sector workers had salaries two to six times higher than those of SOE workers. This ratio corresponds to earnings gaps (in logs) in the range of -0.7 to -1.8. According to Table 7, fewer than 10 percent of SOE workers earn so much less than their private sector counterparts (the earnings gap at the tenth percentile is always higher than –0.7). In statistical terms, these workers represent the "tail" of a distribution, and their estimated earnings gaps are likely to reflect measurement error, rather than dismally low remuneration in the public sector.

A fourth, important conclusion refers to the consistency of the aggregate results obtained with the different econometric models used in this paper. This consistency is highlighted by Figures 1 to 3, which reproduce the distribution of the earnings gaps estimated with all five models. Each of these figures corresponds to one of the three earnings indicators considered. The figures were drawn using an Epanechnikov kernel density function with fifty intervals. Figures 1 to 3 display a similar distribution of earnings gaps, regardless of the econometric technique used.

Last but not least, the results obtained with different econometric models are also consistent at the individual level. In principle, the distribution of the earnings gap could

be similar with all five models, but the place occupied by different SOE workers in these distributions could be dramatically different. For instance, a specific SOE worker could appear to be generously overpaid when using one model, and dismally underpaid when using a different one. The top portions of Tables 7 to 9 show that this is not the case. The correlation coefficient between individual earnings gaps estimated with any two models is always high. It exceeds 0.7 in more than half of the cases.

8. Robustness

The earnings gaps estimated in this paper are presumably robust, in the sense that five different econometric models and three different earnings variables were used in the process, all leading to relatively similar conclusions. However, criticism could be made that these estimates are flawed, as they are all based on self-reported earnings. The under-declaration of earnings is a common bias of household surveys like the VLSS. This bias provides a strong rationale to use the jobs approach when estimating the earnings gap between the public and the private sector. The jobs approach rests on earnings data collected from establishments, hence much less subject to measurement error. Imagine, for instance, that workers systematically report a third of their earnings. In that case, salaries in the "best" companies in Vietnam would be only twice as high as SOE salaries, and not six times higher as Table 1 suggested.

While this criticism is potentially valid, it does not appear to be highly relevant in the case of Vietnam. Based on the VLSS, in 1998 the average annual compensation of an SOE worker in his or her occupation was 8,728 thousand dong. On the other hand, an enterprise survey carried out in 2000 reported an average compensation of 10,270 thousand dong per SOE worker (Belser and Rama, 2001, Table 2). This latter figure being based on firm-level records, it is similar in nature to the information the jobs approach would use. But is it really higher than the one resulting from the VLSS? Between 1998 and 2000 the Vietnamese economy grew at an average rate of roughly five percent per year. If the growth rates of salaries were similar, 8,728 thousand dong in 1998 would become 9,623 thousand dong in 2000. Moreover, the enterprises considered in the above-mentioned survey were large, and it is common for large enterprises to pay higher

wages. Taking into account wage inflation and this size bias, SOE wages do not appear to be under-estimated by the VLSS.

A related concern has to do with the different extent of under-reporting in the public and the private sectors of the economy. Maybe SOE salaries are not under-reported, but the more variable and less well documented earnings of private sector workers are. Systematic under-reporting of private sector earnings would lead to overestimate the earnings gap, and possibly support the conclusion that SOE workers are overpaid, even when this were not true.

One way to address this concern is to estimate consumption gaps, instead of earnings gaps. Surveys like the VLSS have a very detailed consumption module. Respondents are asked in great detail about which items they purchase, how frequently, and at which cost. Such a detailed measurement of consumption is deemed necessary for poverty assessments. But in the present context it can be used to verify whether SOE workers have a higher level of consumption per capita than their private sector counterparts (as the results in the previous sections show) or a lower level (as the jobs approach suggests).

All but one of the econometric models were therefore re-estimated using consumption per capita, instead of earnings, as the dependent variable. The consumption figures were constructed by adding up rice, other food, non-food expenditures (including health and education), depreciation of durable goods and the rental price of housing, using a methodology that is standard in poverty assessments. These values were deflated by rice, other food, and non-food prices, for each region and each month of the year. The resulting total was divided by household size. Because consumption per capita is affected by the size and age composition of a household, as well as by its accumulated assets, additional explanatory variables were needed when re-estimating the models. Those variables are the ones listed in the third data panel in Table 2. Household effects is the one model that could not be re-estimated. This is because, by construction, all of the adult members of a household have the same consumption level. In the absence of variation in the dependent variable within the household, differences in the observable characteristics

of its members (including the sector they work in) would appear to have no consequences.

The estimated consumption gaps are summarized in Table 10. It appears that the consumption per capita of SOE workers would decline by roughly twenty percent if they were to move to the private sector. Only one model (the last one) yields a smaller gap, but it still shows that most SOE workers would be worse off if they had to relinquish their jobs. As before, the correlation between gaps estimated with any two models is high.

9. Conclusion

The workers approach provides reliable estimates of the earnings gap between public and private sector jobs, even in a country where recruitment into the public sector, and separations from it, are strongly influenced by individual characteristics that are not observable. All of the estimates of the average earnings gap fall within a relatively narrow range, regardless of the model used. The overall distribution of earnings gaps at the individual level is similar in all cases. And individual earnings gaps estimated with any two models are highly correlated. Based on the results obtained, SOE workers would earn similar or slightly lower hourly wages in the private sector, but their total earnings would fall by more than 20 percent. It is thus safe to conclude that SOE workers are overpaid, not underpaid, in Vietnam.

If anything, the results in this paper underestimate the extent to which SOE workers are overpaid, because it focuses on measurable earnings and benefits only. Some of the most valuable benefits associated with public sector jobs are very difficult to quantify. They include higher job security, a more generous old-age pension regime, more flexibility, and lower effort levels, among others. None of these benefits was included in the earnings variables used in the analysis. In developing countries, only the "best" private sector jobs offer similar benefits. If SOE workers were to move to the private sector, they would therefore experience a larger loss than measured by the earnings gap only. Studies done for other countries estimate the value of the loss in

intangible benefits at 20 to 50 percent of the SOE salary (see Assaad, 1999, and Chong and Rama, 2001).

The conclusion that SOE workers are substantially overpaid is at odds with the one obtained when using the jobs approach. The latter suggested that SOE workers could earn salaries two to six times higher in the private sector. Admittedly, one could argue about the most accurate econometric model to estimate the earnings gap when using the workers approach. It is also regrettable that some of the estimates of the average earnings gap differ by as much as twenty percentage points. But the range of variation resulting from the workers approach, depending on the model used, is negligible compared to the magnitude of the mistakes the jobs approach would lead to. Governments, and multilateral organizations such as the World Bank, would be ill-advised to recommend pay raises in the public sector based on the jobs approach. The results in this paper show that even a crude version of the workers approach, not correcting the estimated earnings gaps for selection bias, would yield more reliable results.

Of course, it would be inappropriate to conclude that public sector workers are always overpaid. The results in this paper refer to a specific group of public sector workers in a specific country at a specific point in time. The results may not even apply to government employees in Vietnam. In fact, the most important message of this paper refers to the methodology it proposes, not to the results it obtains. This methodology could be easily replicated for other public sector workers, especially in the context of public sector reform. In the case of Vietnam, government administration will be subject to an ambitious reform program over the next ten years. According to this program, "fundamental reforms will be undertaken with regard to the salary of cadres and civil servants, so that these will become the main driving force for the public administration system and be adequate to maintain their life and the life of their families" (Government of Vietnam, 2001, p. 10, official translation). Basing those fundamental reforms on the workers approach, rather than the jobs approach, would be advisable.

In finishing, it is worth making one last point on methodology. All of the models used in this paper are based on a hypothetical comparison of earnings in and out of the public sector. Experimentation would be a much more reliable method to estimate the

earnings gap. If SOE workers could be fired, their subsequent earnings would provide an unbiased measure of the extent to which they were underpaid or overpaid in the public sector. Needless to say, running such an experiment would be inadmissible on moral grounds. But the reform program of Vietnam, aimed at modernizing the state sector, will mimic it in practice.

Over the next few years several hundred thousand SOE workers may lose their jobs as their enterprises change ownership, are restructured or go bankrupt. Most of the job separations will be voluntary, but for some workers there will be no choice (for instance, in the context of liquidation). Earnings gaps estimated along the lines proposed in this paper have been used to design an appropriate compensation package (World Bank, 2001). It is precisely because SOE workers are overpaid that the package needs to be generous. Indeed, there would be no need for compensation if separated workers could earn two to six times higher salaries after separation, as the jobs approach indicates. But information on the exact extent of the subsequent loss in earnings can be used to fine-tune the compensation package. And this information can also be used to assess which of the five models considered in this paper produced more accurate (or less inaccurate) predictions. Such an assessment is the natural follow up of this research.

References

Adamchik, Vera A., and Arjun S. Bedi (2000): "Wage Differentials between the Public and the Private Sector: Evidence from an Economy in Transition", *Labour Economics*, 7(2), p. 203-224, March.

Alderman, Harold, Sudarshan Canagarajah, and Stephen D. Younger (1996): "Consequences of Permanent Layoff from the Civil Service: Results from a Survey of Retrenched Workers in Ghana", in David Lindauer and Barbara Nunberg (eds.): *Rehabilitating Government: Pay and Employment Reform in Africa*, Washington, DC: The World Bank.

Assaad, Ragui (1999): "Matching Severance Payments with Worker Losses in the Egyptian Public Sector", *World Bank Economic Review*, 13(1), p. 117-154, January.

Bales, Sarah (2000): "Vietnam's Labor Situation and Trends – Analysis Based on 1992-93 and 1997-98 Vietnam Living Standards Survey Data", background paper to the *Vietnam Development Report 2000*, Hanoi: The World Bank.

Belser, Patrick, and Martín Rama (2000): "State Ownership and Labor Redundancy: Estimates Based on Enterprise-Level Data from Vietnam", *unpublished manuscript*, Washington, DC: The World Bank.

Chong, Alberto, and Martín Rama (2001): "Do Compensation Packages Need to Be that Generous? Simulations for Government Employees in Guinea-Bissau", in Shantayanan Devarajan, F. Halsey Rogers and Lyn Squire (eds.): *World Bank Economists' Forum*, 1, p. 169-194, Washington, DC: The World Bank.

Filmer, Deon, and David L. Lindauer (2001): "Does Indonesia Have a 'Low Pay' Civil Service?", forthcoming in *Bulletin of Indonesian Economic Studies*.

Glewwe, Paul, and Margaret Grosh (1998): "Data Watch: The World Bank's Living Standards Measurement Study Household Surveys", *Journal of Economic Perspectives*, 12(1), Winter.

Government of Vietnam (2001): "Strategy for Public Administration Reform for the Period 2001-2010", *unpublished manuscript*, Hanoi: Ministry of Planning and Investment.

Haltiwanger, John, and Manisha Singh (1999): "Cross-Country Evidence on Public Sector Retrenchment", *World Bank Economic Review*, 13(1), p. 23-66, January.

Hou, Jack W. (1993): "Public-Private Wage Comparison: A Case Study of Taiwan", *Journal of Asian Economics*, 4(2), p. 347-362, Fall.

Kikeri, Sunita (1997): "Privatization and Labor: What Happens to Workers when Governments Divest?", *World Bank Technical Paper*, 396, Washington, DC: The World Bank.

Lakshmanasamy, T., and S. Ramasamy (1999): "An Econometric Analysis of the Work Choice between Public and Private Sector", *Indian Journal of Labour Economics*, 42(1), p. 71-83, January-March.

Lindauer, David, and Richard H. Sabot (1983): "The Public-Private Wage Differential in a Poor Urban Economy", *Journal of Development Economics*, 12(1-2), p. 137-152, February-April.

Mengistae, Taye (1998): "Wage Rates and Job Queues: Does the Public Sector Overpay in Ethiopia?", *Working Paper Series*, WPS/98-20, Oxford, UK: Centre for the Study of African Economies", December.

MOLISA (1998): *Female Labor Force in Industry in Vietnam in Period of Renovation*, Hanoi: Ministry of Labor Invalids and Social Affairs, Institute for Labor and Social Policy.

Psacharopoulos, George, Jorge Valenzuela and Mary Arends (1996): "Teacher Salaries in Latin America: a Review", *Economics of Education Review*, 15(4), p. 401-406.

Rama, Martín (1999): "Public Sector Downsizing: an Introduction", *World Bank Economic Review*, 13(1), p. 1-22, January.

Rama, Martín (2001): "The Gender Implications of Public Sector Downsizing: the Reform Program of Vietnam", *Policy Research Working Paper*, 2573, Washington, DC: The World Bank.

Rama, Martín, and Donna MacIsaac: "Earnings and Welfare after Downsizing: Central Bank Employees in Ecuador", *World Bank Economic Review*, 13(1), p. 89-116, 1999.

Stelcner, Morton, Jacques van de Gaag and Wim Vijverberg (1989): "A Switching Regression Model of Public-Private Sector Wage Differentials in Peru: 1985-86", *Journal of Human Resources*, 24(3), p. 545-559, Summer.

Terrell, Katherine (1993): "Public-Private Wage Differentials in Haiti: Do Public Servants Earn a Rent?", Journal of Development Economics, 42(2), p. 293-314, December.

van der Gaag, Jacques, and Wim Vijverberg: "A Switching Regression Model for Wage Determinants in the Public and Private Sectors of a Developing Country", *Review of Economics and Statistics*, 70(2), p. 244-252, May.

World Bank (2001): "A Proposed Poverty Reduction Support Credit to the Socialist Republic of Vietnam: Report and Recommendation of the President", *Report No: P-7446-VN*, Washington, DC: The World Bank.

Table 1. Salaries Based on Jobs Approach

	Data from household survey		*Data from "best" employers*	
			Private sector workers	
Job description	*SOE workers*	*Private sector workers*	*At entry level*	*At mid- career*
Cleaners	14,346	n.a.	37,764	44,028
Guards	7,904	14,661	n.a.	n.a.
Drivers	10,670	15,391	n.a.	n.a.
Office workers	10,420	23,623	53,436	62,364

Source: All figures are in thousand dong, at 1998 prices. The first two data columns report authors' calculations based on data from the 1997-98 VLSS. The private sector jobs considered in the second data column are for workers who have a written contract in enterprises with a personnel of 10 or more. The figures in the last two columns are from an unpublished salary review commissioned by the Vietnam office of the United Nations Development Program (UNDP) in 2000. They are adjusted to 1998 prices based on the variation of the exchange rate index.

Table 2. Descriptive Statistics

	SOE workers			*Private sector workers*		
Variable	*Mean*	*Median*	*St. dev.*	*Mean*	*Median*	*St. dev.*
Annual compensation in all jobs	8,970	6,729	9,342	5,494	4,357	5,111
Annual compensation in main job	8,728	6,534	9,198	5,305	4,218	4,996
Hourly compensation in main job	3.86	2.89	3.97	2.91	2.33	2.67
Annual per capita consumption	5,061	4,318	3,395	2,935	2,224	2,539
Schooling (years)	10.52	12.00	3.42	6.80	7.00	3.75
Vocational training (years)	0.69	0.00	1.21	0.14	0.00	0.59
Work experience (years)	17.2	17.0	9.9	17.3	14.0	12.1
Male (yes = 1)	0.56	1.00	0.50	0.63	1.00	0.48
Religion (any religion = 1)	0.21	0.00	0.41	0.37	0.00	0.48
Kinh or Chinese (yes = 1)	0.99	1.00	0.10	0.93	1.00	0.26
Single (yes = 1)	0.33	0.00	0.47	0.49	0.00	0.50
Household head (yes = 1)	0.30	0.00	0.46	0.27	0.00	0.44
Lives in a small town (yes = 1)	0.18	0.00	0.38	0.13	0.00	0.33
Lives in a medium town (yes = 1)	0.25	0.00	0.43	0.09	0.00	0.29
Lives in a major city (yes = 1)	0.27	0.00	0.44	0.20	0.00	0.40
Lives in a delta area (yes = 1)	0.81	1.00	0.39	0.76	1.00	0.43
Household size	5.15	5.00	2.16	5.57	5.00	2.12
Share of children in household (%)	23.5	22.2	19.4	24.9	25.0	20.8
Share of elderly in household (%)	8.53	0.00	14.6	8.50	0.00	14.6
Agricultural land (square meters)	0.13	0.00	0.49	0.20	0.00	0.51
Road in community (yes = 1)	0.97	1.00	0.16	0.85	1.00	0.36
Cottage industry in community (yes =1)	0.68	1.00	0.47	0.64	1.00	0.48
Factory in community (yes = 1)	0.91	1.00	0.29	0.70	1.00	0.46
Number of observations		566			1,970	

Source: Authors' calculations based on data from the 1997-98 VLSS. All monetary figures are in thousand dong.

Table 3. Estimates without Correcting for Selection Bias

	The dependent variable is the log of		
Explanatory variables	*Annual earnings in main job*	*Annual earnings in all jobs*	*Hourly earnings in main job*
Schooling (years)	0.0405 ***	0.0411 ***	0.0428 ***
	(6.677)	(6.455)	(8.437)
Vocational training (years)	-0.0007	0.0004	-0.0010
	(-0.023)	(0.012)	(-0.032)
Work experience (years)	0.0205 ***	0.0210 ***	0.0142 ***
	(3.700)	(3.614)	(3.362)
Work experience squared	-0.0005 ***	-0.0005 ***	-0.0003 ***
	(-5.165)	(-5.231)	(-4.144)
Male (yes = 1)	0.3010 ***	0.3111 ***	0.2768 ***
	(7.788)	(7.960)	(8.865)
Religion (any religion = 1)	-0.0980 *	-0.0942 *	-0.0434
	(-1.715)	(-1.755)	(-1.211)
Kinh or Chinese (yes = 1)	0.0230	0.0525	0.0441
	(0.279)	(0.619)	(0.864)
Single (yes = 1)	-0.0812 *	-0.0878 *	-0.0724 **
	(-1.767)	(-1.893)	(-1.985)
Household head (yes = 1)	-0.0138	-0.0152	-0.0432
	(-0.307)	(-0.329)	(-1.183)
Lives in a small town (yes = 1)	-0.0851	-0.0910	-0.0879 *
	(-1.168)	(-1.277)	(-1.753)
Lives in a medium town (yes = 1)	0.0950	0.0890	0.0559
	(1.016)	(0.999)	(0.841)
Lives in a major city (yes = 1)	0.4351 ***	0.4101 ***	0.2736 ***
	(4.390)	(4.257)	(3.455)
Lives in a delta area (yes = 1)	0.0501	0.0384	0.0453
	(0.787)	(0.618)	(0.969)
Road in community (yes = 1)	0.2062 **	0.1716 *	-0.0154
	(2.190)	(1.855)	(-0.299)
Cottage industry in community (yes =1)	-0.0484	-0.0444	-0.0543
	(-0.740)	(-0.694)	(-1.210)
Factory in community (yes = 1)	0.0529	0.0437	-0.0118
	(0.719)	(0.583)	(-0.279)
Regional dummies	Yes	Yes	Yes
R^2	0.278	0.274	0.250
F-test	17.84	17.93	17.82
Number of observations	1,970	1,970	1,970

Source: Authors' calculations. Refers to workers in the private sector only. Estimated on data from the 1997-98 VLSS, using ordinary least squares weighted by sampling cluster. Values in parenthesis are *t*-statistics. Significant coefficients at the 10, 5 and 1 percent levels are indicated by one, two and three asterisks, respectively.

Table 4. Estimates using the Switching Regression Model

	The dependent variable is the log of		
Explanatory variables	*Annual earnings in main job*	*Annual earnings in all jobs*	*Hourly earnings in main job*
Schooling (years)	0.0323 ***	0.0328 ***	0.0373 ***
	(4.547)	(4.689)	(7.032)
Vocational training (years)	-0.0275	-0.0267	-0.0191
	(-0.922)	(-0.887)	(-0.981)
Work experience (years)	0.0184 ***	0.0189 ***	0.0128 ***
	(3.905)	(4.003)	(3.316)
Work experience squared	-0.0005 ***	-0.0005 ***	-0.0002 ***
	(-6.207)	(-6.542)	(-3.820)
Male (yes = 1)	0.3200 ***	0.3302 ***	0.2895 ***
	(9.012)	(9.382)	(10.298)
Religion (any religion = 1)	-0.0895 **	-0.0855 **	-0.0374
	(-2.464)	(-2.327)	(-1.289)
Kinh or Chinese (yes = 1)	0.0190	0.0482	0.0417
	(0.248)	(0.624)	(0.623)
Single (yes = 1)	-0.0735	-0.0802	-0.0676 *
	(-1.485)	(-1.585)	(-1.714)
Household head (yes = 1)	-0.0005	-0.0015	-0.0345
	(-0.010)	(-0.031)	(-0.928)
Lives in a small town (yes = 1)	-0.0974 *	-0.1032 *	-0.0958 **
	(-1.716)	(-1.775)	(-2.130)
Lives in a medium town (yes = 1)	0.0541	0.0477	0.0290
	(0.769)	(0.681)	(0.557)
Lives in a major city (yes = 1)	0.4334 ***	0.4084 ***	0.2728 ***
	(7.298)	(6.837)	(5.880)
Lives in a delta area (yes = 1)	0.0453	0.0339	0.0423
	(0.998)	(0.740)	(1.182)
Road in community (yes = 1)	0.1951 ***	0.1605 ***	-0.0229
	(3.629)	(3.013)	(-0.446)
Cottage industry in community (yes =1)	-0.0269	-0.0227	-0.0403
	(-0.668)	(-0.564)	(-1.204)
Factory in community (yes = 1)	0.0345	0.0251	-0.0243
	(0.791)	(0.574)	(-0.657)
Regional dummies	Yes	Yes	Yes
Correlation $?^0$ between $?_i$ and $?^0_i$	-0.6097 ***	-0.6287 ***	-0.5754 ***
Correlation $?^S$ between $?_i$ and $?^S_i$	0.2950 **	0.2977	0.2560 *
Log of likelihood function	-3447.8	-3451.3	-2893.0
Number of observations	2,536	2,536	2,536

Source: Authors' calculations. Reported coefficients refer to workers in the private sector only. Estimated on data from the 1997-98 VLSS, using full-information maximum likelihood weighted by sampling cluster. Values in parenthesis are *t*-statistics. Significant coefficients at the 10, 5 and 1 percent levels are indicated by one, two and three asterisks, respectively.

Table 5. Estimates using Individual-Specific Random Effects

	The dependent variable is the log of		
Explanatory variables	*Annual earnings in main job*	*Annual earnings in all jobs*	*Hourly earnings in main job*
Main job is in an SOE (yes = 1)	0.4397 ***	0.4207 ***	0.2147 ***
	(7.420)	(7.143)	(4.690)
Schooling (years)	0.0286 ***	0.0331 ***	0.0270 ***
	(3.484)	(4.060)	(4.250)
Work experience (years)	0.0515 ***	0.0511 ***	0.0391 ***
	(6.431)	(6.418)	(6.308)
Work experience squared	-0.0010 ***	-0.0010 ***	-0.0007 ***
	(-7.049)	(-7.124)	(-6.285)
Male (yes = 1)	0.3308 ***	0.3502 ***	0.2656 ***
	(6.396)	(6.823)	(6.635)
Religion (any relig ion = 1)	-0.0885 *	-0.0847 *	-0.0728 *
	(-1.792)	(-1.724)	(-1.912)
Kinh or Chinese (yes = 1)	0.1259	0.1358	0.0500
	(1.164)	(1.263)	(0.598)
Single (yes = 1)	0.0037	-0.0071	-0.0438
	(0.054)	(-0.103)	(-0.819)
Household head (yes = 1)	0.0148	-0.0031	-0.0409
	(0.246)	(-0.052)	(-0.880)
Lives in a small town (yes = 1)	0.0552	0.0427	0.1096 *
	(0.701)	(0.546)	(1.796)
Lives in a medium town (yes = 1)	0.1211	0.0966	0.0421
	(1.418)	(1.140)	(0.638)
Lives in a major city (yes = 1)	0.3896 ***	0.3789 ***	0.2518 ***
	(4.764)	(4.666)	(3.978)
Lives in a delta area (yes = 1)	0.0281	0.0091	0.0660
	(0.442)	(0.143)	(1.340)
Road in community (yes = 1)	0.0746	0.1016	-0.1041
	(0.887)	(1.215)	(-1.603)
Cottage industry in community (yes =1)	0.0721	0.1074 *	0.0538
	(1.231)	(1.844)	(1.191)
Factory in community (yes = 1)	0.1262 **	0.1081 *	0.0291
	(2.064)	(1.776)	(0.618)
Year 1998 (yes = 1)	0.4402 ***	0.4552 ***	0.4188 ***
	(10.428)	(10.827)	(12.882)
Regional dummies	Yes	Yes	Yes
Overall R^2	0.396	0.402	0.400
Wald $?^2$	641.63	657.22	663.09
Hausman test	19.87	17.55	11.43
Number of observations	1,084	1084	1,084

Source: Authors' calculations. Refers to all workers who had earnings in 1992-93 and 1997-98. Estimated on panel data from the VLSS, using random effects at the individual level. Values in parenthesis are *t*-statistics. Significant coefficients at the 10, 5 and 1 percent levels are indicated by one, two and three asterisks, respectively.

Table 6. Estimates using Household-Specific Fixed Effects

	The dependent variable is the log of		
Explanatory variables	*Annual earnings in main job*	*Annual earnings in all jobs*	*Hourly earnings in main job*
Main job is in an SOE (yes = 1)	0.2078 ***	0.2310 ***	0.0355
	(2.945)	(3.164)	(0.691)
Schooling (years)	0.0228 ***	0.0260 ***	0.0275 ***
	(2.788)	(3.062)	(4.617)
Vocational training (years)	-0.0266	-0.0128	-0.0336
	(-0.833)	(-0.385)	(-1.442)
Work experience (years)	0.0397 ***	0.0406 ***	0.0295 ***
	(6.748)	(6.681)	(6.898)
Work experience squared	-0.0008 ***	-0.0008 ***	-0.0005 ***
	(-7.862)	(-7.913)	(-6.261)
Male (yes = 1)	0.2432 ***	0.2800 ***	0.2250 ***
	(7.178)	(7.987)	(9.128)
Single (yes = 1)	-0.0450	-0.0512	-0.0122
	(-0.717)	(-0.790)	(-0.268)
Household head (yes = 1)	0.1114 **	0.0889	0.0323
	(2.126)	(1.635)	(0.848)
Regional dummies	No	No	No
Overall R^2	0.160	0.176	0.141
F-test	24.01	25.98	29.24
Hausman test	30.47	28.66	35.03
Number of observations	1,549	1,549	1,549

Source: Authors' calculations. Refers to all workers in households with at least two wage earners. Estimated on data from the 1997-98 VLSS, using fixed effects at the household level. Values in parenthesis are *t*-statistics. Significant coefficients at the 10, 5 and 1 percent levels are indicated by one, two and three asterisks, respectively.

Table 7. Summary Statistics for Gap in Annual Earnings in Main Job

	Gap between SOEs and private sector				
	R1	*R2*	*R3*	*R4*	*R5*
R1	1.0000	0.9852	0.8854	0.5563	0.7420
R2		1.00000	0.8541	0.5685	0.7295
R3			1.0000	0.3677	0.6597
R4				1.0000	0.4407
R5					1.0000
Mean	0.3269	0.1828	0.4431	0.2078	0.1714
Median	0.3198	0.1736	0.4176	0.2195	0.1907
Standard deviation	0.5687	0.5664	0.3754	0.3662	0.7299
10th %	-0.3203	-0.4620	0.0201	-0.2363	-0.6912
25th %	0.0048	-0.1500	0.2074	0.0346	-0.2724
75th %	0.6849	0.5233	0.6814	0.3767	0.6294
90th %	0.9929	0.8264	0.8830	0.6561	0.9982
Observations	566	566	143	317	368

Source: Authors' calculations.

Table 8. Summary Statistics for Gap in Annual Earnings in All Jobs

	Gap between SOEs and private sector				
	R1	*R2*	*R3*	*R4*	*R5*
R1	1.0000	0.9836	0.8866	0.5726	0.7379
R2		1.0000	0.8521	0.5829	0.7241
R3			1.0000	0.4207	0.6496
R4				1.0000	0.4405
R5					1.0000
Mean	0.3271	0.1748	0.4200	0.2310	0.1675
Median	0.3109	0.1820	0.4026	0.2403	0.1973
Standard deviation	0.5635	0.5629	0.3738	0.3711	0.7291
10th %	-0.3364	-0.5023	-0.0046	-0.2019	-0.6377
25th %	-0.0118	-0.1591	0.1795	0.0406	-0.2631
75th %	0.6735	0.5151	0.6631	0.4213	0.5829
90th %	0.9753	0.8280	0.8628	0.6858	1.0292
Observations	566	566	143	317	368

Source: Authors' calculations.

Table 9. Summary Statistics for Gap in Hourly Earnings in Main Job

	Gap between SOEs and private sector				
	R1	*R2*	*R3*	*R4*	*R5*
R1	1.0000	0.9830	0.8887	0.5374	0.7292
R2		1.0000	0.8533	0.5525	0.7256
R3			1.0000	0.4910	0.4656
R4				1.0000	0.5186
R5					1.0000
Mean	0.0996	-0.0605	0.1906	0.0355	-0.0467
Median	0.0942	-0.0616	0.2181	0.0453	-0.0577
Standard deviation	0.5357	0.5308	0.3363	0.3403	0.7356
10th %	-0.5482	-0.7070	-0.2149	-0.4042	-0.9087
25th %	-0.2239	-0.3889	-0.0502	-0.1279	-0.4708
75th %	0.4045	0.2522	0.4002	0.2247	0.4276
90th %	0.7288	0.5362	0.5701	0.4578	0.8532
Observations	566	566	143	317	368

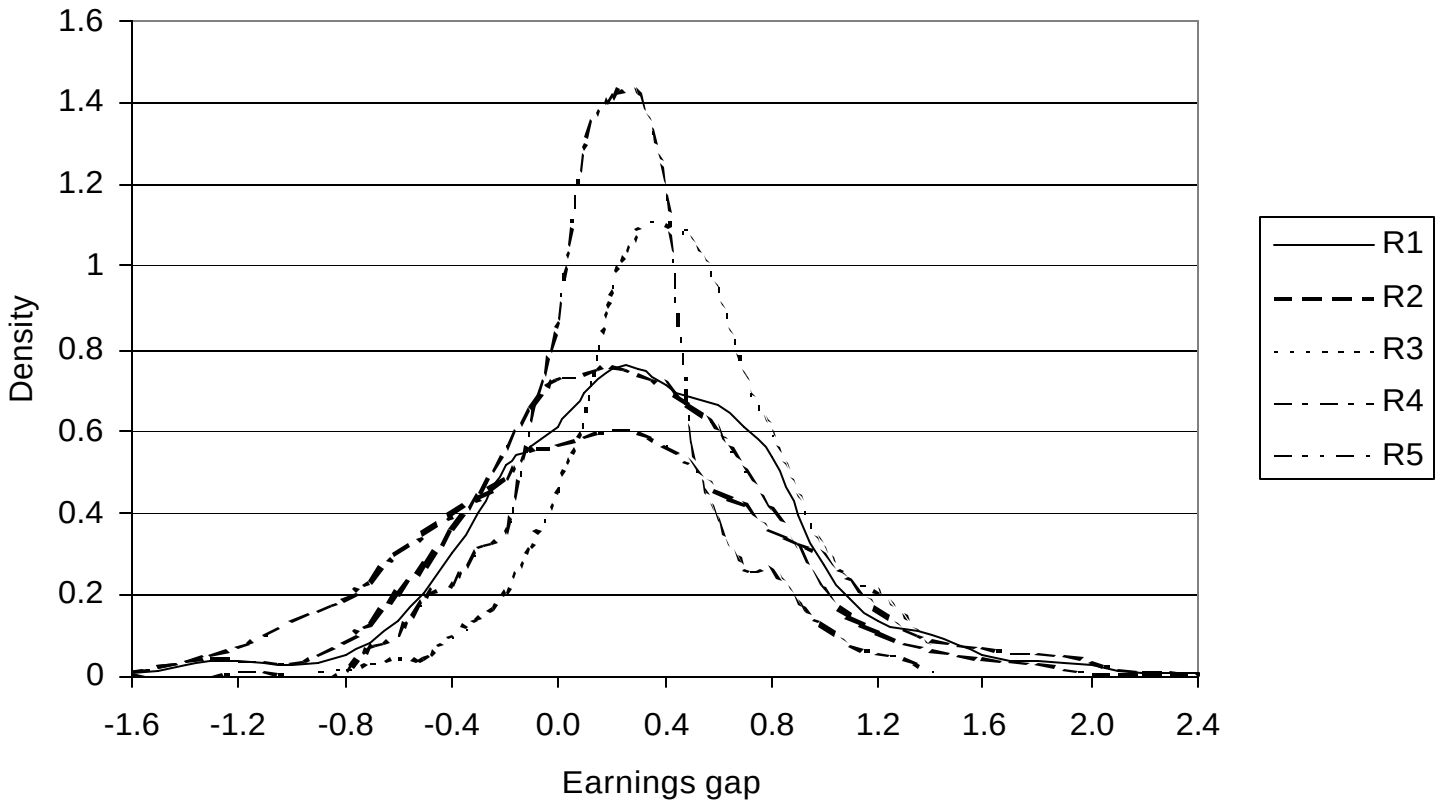
Source: Authors' calculations.

Table 10. Summary Statistics for Gap in Consumption per Capita

	Gap between SOEs and private sector				
	R1	*R2*	*R3*	*R4*	*R5*
R1	1.0000	0.9957	0.6940	n.a.	0.7077
R2		1.0000	0.6825	n.a.	0.7178
R3			1.0000	n.a.	0.5440
R4				1.0000	n.a.
R5					1.0000
Mean	0.2140	0.1880	0.2490	n.a.	0.0753
Median	0.2052	0.1699	0.2404	n.a.	0.0643
Standard deviation	0.4259	0.4284	0.2305	n.a.	0.5611
10th %	-0.3052	-0.3494	-0.0320	n.a.	-0.5938
25th %	-0.0642	-0.0747	0.0536	n.a.	-0.3215
75th %	0.4704	0.4487	0.4094	n.a.	0.4323
90th %	0.7607	0.7179	0.5515	n.a.	0.7203
Observations	566	566	143	0	368

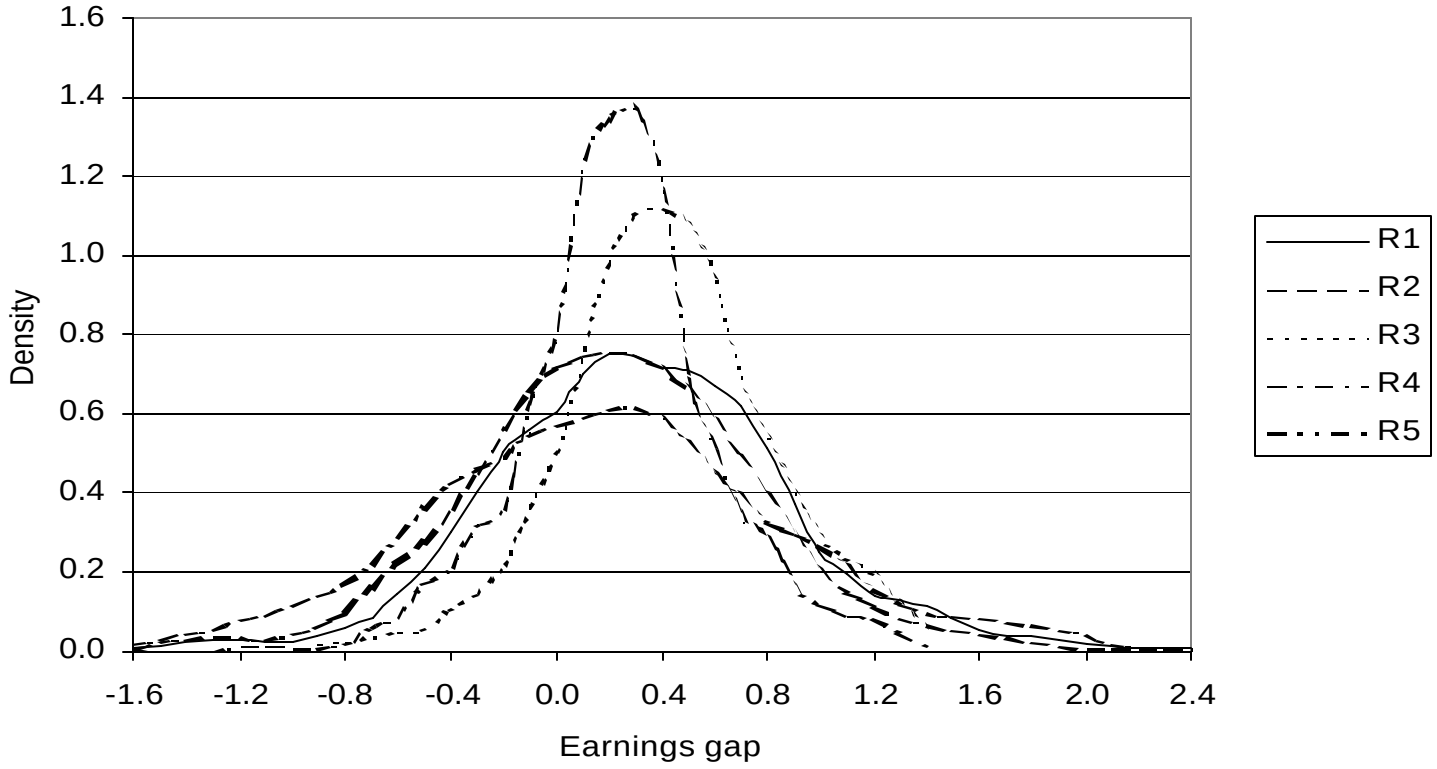
Source: Authors' calculations.

Figure 1. The Distribution of the Gap in Annual Earnings in Main Job



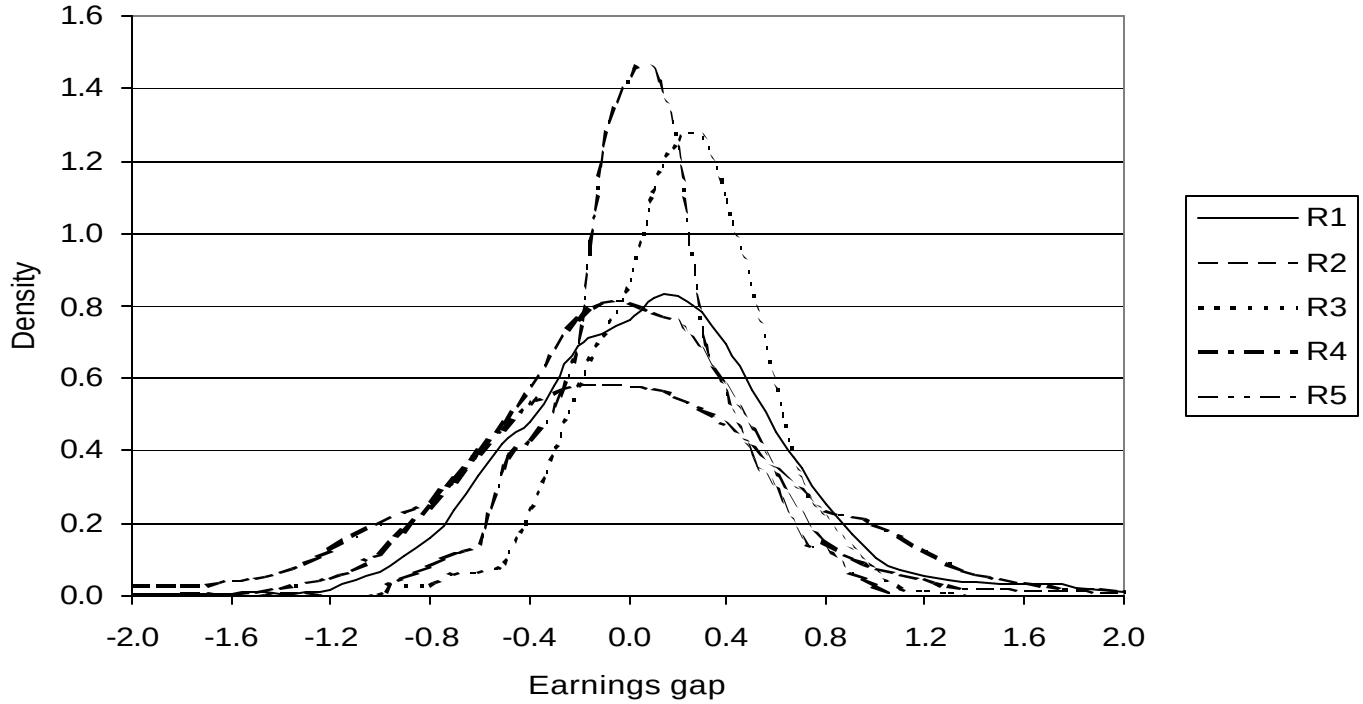
Source: Authors' calculations.

Figure 2. The Distribution of the Gap in Annual Earnings in All Jobs



Source: Authors' calculations.

Figure 3. The Distribution of the Gap in Hourly Earnings in Main Job



Source: Authors' calculations.