SHOULD WORKERS CARE ABOUT FIRM SIZE?

ANA FERRER† AND STÉPHANIE LLUIS‡

†Assistant professor, University of Calgary. ‡Assistant professor, University of Waterloo. The authors thank seminar participants at the University of British Columbia for valuable comments on earlier versions of this paper. They are also grateful to Statistics Canada and the SSHRC for providing access to the SLID data. The data used in this study was provided under contract with Statistics Canada and cannot be released by the authors. Individuals wishing to obtain access to these data should contact the authors or the RDC directly at http://www.statcan.gc.ca/english/rdc/whatdata.htm.
SHOULD WORKERS CARE ABOUT FIRM SIZE?

Abstract

In this paper, we analyze how firms of different sizes reward measured skills and unmeasured ability. Our empirical methodology, based on non linear instrumental variable estimations, permits a direct estimation of the returns to unmeasured ability by firm size. Using data from the SLID, we find that the returns to unmeasured ability are significantly higher in medium-sized firms relative to smaller firms. Also, the returns to unmeasured ability are significantly lower in larger firms relative to medium firms. There seems to be a firm size threshold above which larger firms are not able to identify and reward ability. When firms become “too large”, monitoring costs may prevent firms from rewarding ability directly through wages.
1 Introduction

The existence of firm-size wage differentials, the observation that firms of different sizes pay different wages for observationally equivalent workers, is a widely documented fact in the empirical literature on wage determination. Studies generally find that none of the explanatory variables used in the wage equation can fully explain the size-wage gap. As a result, the literature has concluded that unmeasured factors in the error term explain a substantial portion of the size-wage gap. While the theoretical literature has offered several channels through which unobserved factors might interact with firm size to produce a wage gap, the empirical literature has yet to address the question of how or why unobserved factors might be rewarded differently in firms of different sizes.

In this paper, we reexamine the relationship between firm size and wage outcomes by estimating and testing for differences in the returns to measured skills and especially unmeasured ability by firm size. Our analysis is centered around two well established hypotheses that imply that firms of different sizes will reward unobserved skills differently: i) an ability sorting hypothesis, which states that high ability workers have a comparative advantage in large firms, and ii) a screening hypothesis, according to which large firms find it more costly to monitor aspects of unobserved ability. We use an analytical framework based on a non-random selection model that integrates the different predictions of ability sorting and job screening and apply non linear instrumental variables techniques to longitudinal data on wages and employer size. We find that returns to unmeasured ability initially grow with firm size, but diminish once a certain size is reached. This is a novel result in the literature that

\[1\]Even in the most recent empirical literature which emphasizes issues such as on-the-job-search (Winter-Ebmer and Zweimuller (1999)) or use better information on the employer side using matched employer-employee data (Troske (1999)), the size-wage premium cannot be fully explained by any of the factors analyzed.
analyzes the firm size premium.

The estimated size-wage premium is about 15% in the United States and 10% in Canada. Empirical studies investigating the source of this wage differential have analyzed the explanatory power of various factors related to worker and firm characteristics such as education, unionization and industry type. These studies have been limited in two ways. First, they do not allow for the possibility that human capital attributes may not be equally valued in large and small firms. By doing so, they restrict the effect of firm size on wages to being only a shift parameter. Second, the use of fixed-effect estimations allows one to eliminate the bias caused by the ability term (assuming it is time invariant and equally rewarded in large and small firms) to estimate the effects of other parameters. On the other hand, it does not allow an estimation of ability effects. Moreover, if large and small firms have different wage policies concerning their treatment of measured and unmeasured skills, then the fixed-effect method does not eliminate the effect of unmeasured ability on wage outcomes.

The analysis in this paper estimates the returns to measured and unmeasured skills separately, which enables us to test the relevance of two main explanations proposed in the literature addressing worker non random selection into firms of different sizes, namely ability sorting and job screening. Each argument provides different implications for the importance of the returns to unmeasured skills or ability. According to the ability sorting argument, large firms attract high ability workers because they need better workers as the execution of the different production processes in large

---

2Since studies vary in the information used for firm size categorizations, these numbers are used only for illustrative purpose to show the extent of the wage premium associated with working in a large firm. They correspond to an average across studies of the estimates associated with the highest size category available. The coefficients come from a regression of the wages on firm size dummies controlling for major worker and firm characteristics.
firms is more complex than in small firms.\(^3\) In addition, large firms have more ways to attract better workers than small firms by providing promotion opportunities, training and career development.\(^4\) As a result, returns to unmeasured skills or ability should be greater in large firms than in small ones. In the job screening hypothesis, larger firms have higher monitoring costs than small firms which leads large firms to value observable skills (like education and experience) more and value unobservable skills less compared to small firms.\(^5\)

The method we use relies on a panel data estimator based on the Generalized Method of Moments. It is appropriate for the analysis of multi-sector models of wage determination in which the individual-specific component of the error term is differentially rewarded and selection comes from both sides of the market. The method consists in quasi-differencing the wage equation which involves isolating the individual-specific component in the error term and using instrumental variables estimations to obtain consistent estimates of the returns to measured skills and ability.\(^6\) This type of analysis requires panel data where a sufficiently large number of individuals can be observed for at least three consecutive time periods. Our analysis is applied to longitudinal data from the Canadian Survey of Labour and Income Dynamics (SLID) conducted over the period 1993 to 1998. This household survey provides an extensive set of individual, job and firm characteristics and its panel aspect con-

\(^3\)See Oi (1983) and Idson and Oi (1999) for a theoretical and empirical analysis of the productivity explanation of the size-wage gap.

\(^4\)Agell (2003) conducts an empirical analysis of the characteristics of the compensation policies of establishments by firm size.

\(^5\)See Garen (1985) for a theoretical model and an empirical test supporting this hypothesis. Barron, Black and Loewenstein (1987) and Hu (2003) present evidence of a related implication of the monitoring costs hypothesis. Both empirical studies present findings that are consistent with the idea that large firms use pre-hiring screening practices to reduce monitoring costs associated with on-the-job screening.

\(^6\)Lemieux (1998) applies the method to analyze the effects of union sector choice on wages. Gibbons, Katz, Lemieux and Parent (2005) use this method to analyze inter-occupation and inter-industry wage differentials. Lluis (2005) uses the method to analyze the wage dynamics associated with worker choices of job ranks within a company.
stitutes a unique source of information for an analysis of the dynamic interactions of workers, firm size, and wages.

We start the analysis by replicating the findings of the literature on the size-wage gap in order to validate our data. We proceed to estimate the differential returns to measured skills and unmeasured ability by firm size. We find the data reflect elements of both the ability sorting and job screening stories. The returns to unmeasured ability are significantly higher in medium-sized firms relative to smaller firms. Also, the returns to unmeasured ability are significantly lower in larger firms relative to medium-sized firms. There seems to be a firm size threshold above which larger firms are not able to identify and reward ability. When firms become “too large”, monitoring costs may prevent firms from rewarding ability directly through wages.

Our baseline estimations suffer from two main problems: i) Firm size effects and the returns to ability and skills by firm size are identified through changes in firm size categories for workers changing to a larger or smaller firm. A change in firm size category is also possible without a change of firm if a firm experience substantial hiring or layoffs. To exclude this possibility, we performed the estimations over the sub-sample of involuntary job changers for which a change of firm is more likely to have occurred as well. ii) In the type of non random selection model we use, unmeasured ability is correlated with the choice of firm size but it is also likely the case that ability affects changes in firm size. The statistical model we use can be easily extended to introduce endogenous worker mobility by using appropriate instruments. Furthermore, we also checked for the importance of mis-classification errors in reported firm size and did not find evidence of it. Finally we used establishment size instead of firm size as the measure of employer size to compare possible differences in the results. Our main findings are robust to the various alternative analyses we conducted.
The importance of worker self-selection in the analysis of firm size on wages has been previously analyzed by Idson and Feaster (1990) using cross-sectional data from the 1979 May CPS and Lluis (2004) comparing the size-wage structure in the U.S. and Canada with data from the U.S. Current Population Survey and the Canadian Labour Force Survey for the year 1998. In both of these studies, the method used to characterize non random selection is the Heckman two-step method. Although this method provides evidence of the presence of workers non random selection, it does not provide any quantitative information on the importance of unmeasured ability in the selection mechanism. Moreover, the selection model is one-sided in that it characterizes workers’ decisions to join a firm but does not address firms’ selection decisions. However, the overall effect of firm size on wages depends on both the types of workers that tend to choose to work in a large or a small firm, as well as on the effect of firm size on the wages of different types of workers. The present analysis, in contrast, estimates and tests a wage equation with differential returns to measured and unmeasured skills which allows us to further investigate the consequences of the non random sorting of workers into firms of different sizes and the possibility of differential wage policies by firm size.

Recent studies have analyzed the relationship between firm size and wages using matched employer-employee data. Troske (1999), reaching the same conclusions as the preceding literature, finds that a significant size-wage premium remains even after controlling for the usual suspects and concludes that it is due to unmeasured workforce quality. Abowd, Kramarz and Margolis (1999), Abowd, Finer and Kramarz (1999) and Abowd and Kramarz (2000) estimate a wage equation that includes person-specific and firm-specific unobserved components using longitudinal data on workers and firms from France and from the state of Washington respectively. They find that in both countries, firm-size effects are explained more by firm heterogeneity than
individual heterogeneity. In the framework we use, individual and firm size effects are not treated independently in the wage equation but interacted to illustrate and test for their interdependence in the understanding of the size-wage premium.

2 Analytical Framework

In its attempt to explain the size-wage gap, the empirical literature on the effect of firm size on wages has not considered the possibility that large and small firms may have different wage policies in terms of the rewards to worker skills. Differences in the returns to measured skills and ability and a differential wage structure by firm size can arise for mainly two reasons: ability sorting based on workers comparative advantage into larger firms and differential job screening policies by firm size. Each mechanism leads to different implications in terms of the returns to measured and unmeasured skills depending upon the assumptions made on firms production and monitoring technologies. Ability sorting assumes that skills (measured skills and unmeasured ability) are not identically productive across firms of different sizes and in particular, high ability workers have a comparative advantage in large firms. As a result, the returns to ability should be greater in large firms. Furthermore, the size-wage premium observed in the literature is predicted to disappear as one takes into account and estimates the returns to unmeasured ability. Job screening assumes that monitoring costs increase with the size of the firm so that large firms put less weight on unmeasured ability (and more weight on skills that can be measured with accuracy) relative to small firms as imprecision about the evaluation of ability increases with firm size. Accordingly, the returns to unmeasured ability should be smaller in larger firms. In this case also, the size-wage premium found in the literature results from large firms’ decisions to offer an efficiency wage relative to small firms to attract
high ability workers. This section integrates these two approaches into a statistical model of wage determination and presents the econometric method used to estimate the returns to ability and evaluate the relative importance of workers ability sorting versus firms job screening in explaining the relationship between firm size and wages.

2.1 Implications of Ability Sorting and Job Screening in Wage Policy by Firm Size

The ability sorting argument follows the literature on job assignment and workers comparative advantage first proposed by Roy (1951) and later formalized by Heckman and Honore (1990).\(^7\) The main assumption in this framework is that different skills (measured and unmeasured) are not equally productive across sectors and in the present context, across firms of different sizes. Utility maximizing workers, in this framework, choose the size of employer for which their abilities are best suited. In particular, high ability workers are better suited in larger firms. As a result, the returns to ability are greater in large firms compared to small firms.

Job screening and its implications for the analysis of the size-wage relationship was theoretically developed by Garen (1985). The model relies on the assumption that monitoring/evaluation costs increase with firm size. Larger firms acquire less accurate information about the abilities of their workers than small firms do, and as a result rely less heavily on their own evaluation of ability and more on other indicators of ability such as schooling and experience.

The implications of ability sorting and job screening in terms of the returns to measured skills and unmeasured ability can be described within a single wage equation

\(^7\)See also Neal and Rosen (2000) for more details on selection models and their implications for the earnings distribution.
framework in the following way. Assume for simplicity of exposition that there are two types of firms which differ by size. Firms are indexed by \( j = S \) (small), \( L \) (large). The only input is labor, given in efficiency units per worker. Workers, indexed by \( i = 1,\ldots,N \), are characterized by a vector of productive skills, \((SK_i, \theta_i)\), where \( SK_i \) denotes the observed skills of worker \( i \) and \( \theta_i \) represents traits that are unmeasured by the econometrician (these could include innate ability, initiative, ambition, loyalty).

The idea of different wage policies by firm size can be summarized in the following wage equation:

\[
    w_{ijt} = \alpha_j + \beta_j SK_{it} + \lambda_j \theta_i, \ j = S, L
\]

where \((\beta_j, \lambda_j)\) is the vector of rewards to measured skills and unmeasured ability respectively and \( \alpha_j \) is interpreted as the wage in large and small firms for workers with zero or no ability.\(^8\)

In the ability sorting model, it is assumed that ability is perfectly observed by firms and that workers can be rewarded accordingly.\(^9\) This implies that \( \lambda_L > \lambda_S \) for the returns to unmeasured ability. Since in this model also, measured skills and ability are expected to be positively correlated, the fact that \( \lambda_L > \lambda_S \) implies that \( \beta_L > \beta_S \).\(^10\)

As for the intercepts, the sorting model predicts that \( \alpha_S > \alpha_L \) as pay in small firms is

---

\(^8\)We do not consider the possibility of an absolute advantage in the wage equation as our main focus for the analysis of the relative importance of screening and sorting is the presence of worker comparative advantage. In contrast to comparative advantage, absolute advantage is rewarded identically in each sector (in our case, firm of different sizes). Given that it is unmeasured by the econometrician (assuming it is observed by the market), it would be part of the error term in the wage equation.

\(^9\)For simplicity of exposition, we consider that ability is perfectly observed and focus on the implications of ability sorting and job screening in terms of the differential returns to measured skills and ability by firm size as a result of non random selection of workers into firms of different sizes. We also consider the possibility of endogenous worker mobility. Because the introduction of endogenous mobility does not affect the present discussion of the returns to skills and ability by firm size but mainly affects the choice of instruments, we discuss it in the estimation section.

\(^10\)Define for example \( \beta_j = k \ast \lambda_j \) and assume \( 0 < k \leq 1 \). Gibbons, Katz, Lemieux and Parent (2005) find evidence that across occupations and industries, the estimated returns to unobserved ability are proportional to the returns to measured skills and that the estimated proportionality factor is significantly less than one.
assumed to be less sensitive to unmeasured ability at all levels including the minimum level. Figure 1 illustrates the resulting optimal sorting based on unmeasured ability implied by the ability sorting model.

In the job screening model it is assumed that firms evaluate ability with their own screen which is an imperfect indicator of ability and that imprecision increases with firm size. Greater imprecision in the measurement of ability experienced by larger firms leads them to make pay decisions that rely less heavily on their own evaluation of workers ability than do small firms, and to put more weight on other indicators of ability, such as schooling and experience, which are measured precisely. In this framework, the effect of measured and unmeasured skills on wage outcomes will vary by firm size depending on the degree of accuracy with which worker skills can be evaluated. In particular, the size-wage structure will be characterized by greater returns to measured skills ($\beta_L > \beta_S$), and smaller returns to unmeasured skills ($\lambda_L < \lambda_S$) in larger firms. Differences in the returns to ability (measured and unmeasured) are a direct consequence of the imperfect observation of ability in large firms. In terms of the $\alpha_j$ parameters, the job screening hypothesis implies a larger intercept in large firms ($\alpha_L > \alpha_S$). The presence of an intercept effect $\alpha$ is expected in this model as the lack of accurate information about ability may lead large firms to offer greater wage levels than smaller firms to be able to attract higher

\footnote{Note that in the empirical analysis, interpretation of the estimated alpha parameters is difficult as it depends on the value of average ability $\theta$ which is unmeasured. We should however expect to find a drop in the firm-size effect $\alpha_L - \alpha_S$ relative to the OLS estimate found in the literature due to the omitted variable bias implied by the omission of unmeasured ability in the OLS estimations.}

\footnote{Garen’s model goes beyond the simple characterization of the compensation policies of firms of different sizes in that it endogenizes schooling decisions. This part of the model is not detailed in the present paper as it is not directly relevant to the empirical question of the returns to ability in the size-wage relationship.}

\footnote{See Garen (1985) for more details on the development of these predictions. The measurement issue and the implication for the coefficients associated with skills and ability can also be discussed within the context of an attenuation bias in a typical measurement error problem.}

\footnote{Note that the monitoring cost (or job screening) model presented here can also be viewed as a multi-factor ability sorting model in which sorting is done on both measured and unmeasured ability.
ability workers. Indeed, the literature interpreting the existence of a size-wage gap has referred to the idea that large firms would pay an efficiency wage (Krueger and Summers (1988)) for dealing with supervision and monitoring issues. The implied sorting based on unmeasured ability in the job screening case is illustrated in figure 2.

Although ability sorting and job screening both predict differential returns to measured skills and unmeasured ability by firm size, each mechanism offers opposite predictions for the direction of the difference in returns. If ability sorting based on comparative advantage is most important, then workers selection into large firms is driven mainly by unmeasured ability and the size-wage gap previously estimated in the literature reflects unmeasured ability effects previously omitted when using OLS (or discarded when using fixed-effects) estimations. If the returns to unmeasured ability are smaller in large firms and at the same time, measured ability is more rewarded in larger firms, this suggests the relatively greater importance of the monitoring and screening hypothesis for explaining workers sorting into firms of different sizes.

Overall these predictions allow us to conduct different tests to emphasize the relative importance of each hypothesis. In particular, we can use one-sided tests of equality of the returns to unmeasured skills such that rejecting the null $H_0: \lambda_L = \lambda_S$ would show evidence in favor of the monitoring ($H_A: \lambda_L < \lambda_S$) or the sorting hypotheses ($H_A: \lambda_L > \lambda_S$). It is also possible to test the monitoring cost hypothesis based on the simultaneous inequalities $\lambda_L < \lambda_S$ and $\beta_L > \beta_S$ using a nonlinear one-sided Wald test of equality of the ratios of the returns to measured skills and unobserved ability. Defining $H_0: \frac{\beta_L}{\lambda_L} = \frac{\beta_S}{\lambda_S}$, against $H_A: \frac{\beta_L}{\lambda_L} > \frac{\beta_S}{\lambda_S}$, the rejection of the null suggests that job screening is important in the size wage structure.  

---

15See Garen (1985) for the formal derivation of the above parameter restrictions.
16For an upper one-tailed test with significance level $\alpha$ in a regression model with k explanatory variables and N observations, the null hypothesis of equality is rejected if the value of the statistic
We will concentrate mainly on the slope parameters, the $\lambda$ and $\beta$ coefficients for drawing conclusions about ability sorting and job screening. We will interpret the finding of significant intercepts for larger firms (even after controlling for measured skills and unmeasured ability) as secondary evidence in favor of the monitoring cost hypothesis.

### 2.2 Estimation Method

This section describes the estimation method used to estimate wage equation (1). The empirical version of (1) can be written as follows:

$$ w_{ijt} = \sum_{j=1}^{J} D_{ijt} \alpha_j + \sum_{j=1}^{J} D_{ijt} SK_{it} \beta_j + \sum_{j=1}^{J} D_{ijt} \lambda_j \theta_i + \mu_{it} \tag{2} $$

where $J$ is a discrete variable describing different categories of firm size, $D_{ijt}$ are dummy variables indicating whether individual $i$ is in firm of size $J$ at time $t$, and $\mu_{it}$ is a random noise assumed to be iid.

Estimating equation (2) with OLS would give inconsistent estimates. Indeed, the hypothesis of differential returns to unmeasured ability (as a result of workers comparative advantage or firm screening) implies that firm size assignment is endogenous and unmeasured ability $\theta_i$ is correlated with firm size. Since this framework defines differential returns to unmeasured ability by firm size, ability cannot be eliminated by first-differencing (2) because it is interacted with the firm size dummies $D_{ijt}$.

It is possible to solve this problem by quasi-differencing (2).\textsuperscript{17} To do so we isolate unmeasured ability $\theta_i$ in equation (2) as follows:

$\text{is greater than the critical value } F(\alpha,k-1,N-k).$

\textsuperscript{17}This technique has been used previously by Lemieux (1998) in the case where the returns to a time-invariant unobserved characteristic is different in the union and non-union sector.
\[ \theta_i = \frac{w_{ijt} - \sum_j^J D_{ijt}\alpha_j - \sum_j^J D_{ijt}SK_{it}\beta_j - \mu_{it}}{\sum_j^J D_{ijt}\lambda_j} \]  

Innate ability is time-invariant so its period t version is also equal to its lagged version. To obtain the quasi-difference, the first lag version of (3) is substituted back into equation (2):

\[ w_{ijt} = \sum_{j=1}^J D_{ijt}\alpha_j + \sum_{j=1}^J D_{ijt}SK_{it}\beta_j + \frac{\sum_j^J D_{ijt}\lambda_j}{\sum_j^J D_{ijt-1}\lambda_j} \left[ w_{ijt-1} - \sum_{j=1}^J D_{ijt-1}\alpha_j - \sum_{j=1}^J D_{ijt-1}SK_{it-1}\beta_j \right] + \epsilon_{it} \]  

where \( \epsilon_{it} = \mu_{it} - \frac{\sum_j^J D_{ijt}\lambda_j}{\sum_j^J D_{ijt-1}\lambda_j} \mu_{it-1} \)  

This equation cannot be estimated using non-linear least squares because \( w_{ijt-1} \) is correlated with \( \mu_{it-1} \). Nevertheless, consistent estimates can be used by choosing appropriate instruments for \( w_{ijt-1} \).

The set of instruments, \( Z_i \), has to satisfy the following condition:

\[ E(\epsilon_{it}Z_i) = 0 \]  

The objective is then to minimize the following quadratic form:

\[ \min_{\gamma} \epsilon(\gamma)'Z(Z'\Omega Z)^{-1}Z'\epsilon(\gamma) \]
where $Z'\Omega Z$ is the covariance matrix of the vector of moments $Z'e(\gamma)$, $\Omega$ is the covariance matrix of the error term $e_{it}$ and $\gamma$ is the vector of parameters.

We could use non linear two-stage least squares to estimate equation (2). The estimator would be consistent but not efficient as it assumes that $\Omega$ is homoscedastic. We will use instead the GMM estimator which is consistent and efficient, allowing a general form for $\Omega$ to correct for potential heteroscedasticity and autocorrelation in the error term.  

2.3 Instruments

For the choice of instruments, we use the same strategy as Lemieux (1998) which consists in using the identification assumption for estimation of panel data equations that imposes strict exogeneity of right-hand side variables. More formally:

$$E(\mu_{it}/SK_{i1}...SK_{iT},D_{i1}...D_{iT},\theta_i) = 0 \quad (8)$$

Equation (8) states that conditional on unmeasured innate ability, individual characteristics and firm size assignments each period are uncorrelated with the error term in the wage equation (2). Hence it provides a set of potentially valid instruments with the property that they are not correlated with the $\mu$ terms in the error term $e$ from equation (5).

The comparative advantage and screening assumptions both predict that wages relate to firm size assignment through the interaction of firm size and measured skills.

18This involves a two-step procedure in which the first step estimates the matrix of variance-covariance of the error term $\Omega$ by estimating (7) with $\Omega = I$ (where $I$ is the identity matrix). Note also that for the parameters to be identified, the optimization problem (7) needs the constraint that unmeasured ability $\theta_i$ in the error term of equation (2) sums to zero overall individuals and time periods. Explanation and a proof of the necessity of this constraint is given in Lemieux (1998).
and ability. Therefore we choose to instrument workers’ previous period wage with previous period skills interacted with firm size \( (D_{ijt-1} * SK_{it-1}) \). We also use the history of firm size assignments (the interaction between \( D_{ijt-1} \) and \( D_{ijt} \)). Both instruments help predict lagged wages given that wages are a function of firm size assignments and the interaction between firm size and skills.

### 2.4 Endogenous Mobility

Although quasi-differencing and instrumenting lagged wage is the appropriate technique for addressing the problem of non random assignment of workers into firms of different sizes, the method assumes that after controlling for ability, workers’ choice of firm size is exogenous. Put differently, the above exogeneity assumption (8) rules out the possibility of endogenous firm size switching. With endogenous mobility of workers across firms of different sizes, for example due to learning about ability, temporary productivity shocks would be correlated with the choice of firm size, \( D_{ijt} \). As individuals learn that their ability would match better a larger or smaller firm, they choose to move to a firm of different size.

The present statistical framework can be extended to introduce endogenous job switching based on learning about ability. In this case, agents do not observe ability perfectly. The wage equation depends on expected ability \( \theta_{it}^e = E[\theta_i|z_0, ...z_{t-1}] \) which is a function of beliefs about ability, updated every period from observing the worker’s previous period output (\( z_0, ...z_{t-1} \) is the sequence of imperfect signals of the worker’s ability).

Quasi-differencing is still possible and leads to a similar non linear wage equation as equation (4),\(^\text{19}\) which can be estimated with the GMM estimator. The main

\(^{19}\)See Gibbons, Katz, Lemieux and Parent (2005) and Lluis (2005) for more details about the statistical model when workers mobility is endogenous due to learning effects.
difference when mobility is endogenous is that firm size $D_{ijt}$ needs to be instrumented in addition to instrumenting lagged wages. Also the error term of the wage equation (5) contains the current period random shock which affects the changes in expected ability (the error term of the martingale process) and is correlated with $D_{ijt}$.

The validity of instruments in the case of endogenous mobility relies on the martingale assumption for the evolution of beliefs about ability in addition to the exogeneity condition (8). Given that individuals and firms have rational expectations, only a random productivity shock at time t can change beliefs about ability between t-1 and t (resulting in mobility to a larger or smaller firm at time t). Therefore, the history of previous period choices of firm size (the interaction between $D_{ijt-2}$ and $D_{ijt-1}$) is uncorrelated with the current period changes in expected ability which enters the new residual equation (5). Also given (8), $D_{ijt-2} * D_{ijt-1}$ is also uncorrelated with the error term $\mu$ of the wage equation.

These new instruments should help predict current choice of firm size and lagged wages. The explanation is based on the fact that both mechanisms rely on the assumption that ability (and measured skills) is the main driver of workers’ choice of firm size. The term $D_{ijt-2} * D_{ijt-1}$ reflects the previous period value of expected ability (workers with higher expected ability are more likely to be in larger (smaller) firms if the comparative advantage (job screening) explanation prevails). Under rational expectations, the best prediction of current period expected ability (affecting current choice of firm size) is previous period expected ability. As a result, $D_{ijt-2} * D_{ijt-1}$ should be a good predictor of current period choice of firm size. Similarly, it should help predict lagged wages, a function of previous period expected ability. We will also use lagged versions of interactions between skills and firm size choices. In summary, the main instruments for lagged wages and for $D_{ijt}$ will be based on the following set
of variables: $D_{ijt-2} \times D_{ijt-1}, SK_{ijt} \times D_{ijt-2}, SK_{ijt-1} \times D_{ijt-1}$.\textsuperscript{20}

3 Data and Preliminary Analysis

In this section we describe the data and present an analysis of the size-wage gap in the same spirit as the empirical literature on firm size and wage outcomes.

3.1 Description of the Data

Our data comes from the Survey of Labour and Income Dynamics (SLID). Statistics Canada collects an extensive set of individual and job characteristics for a group of workers, who are followed over a six year period. We use two panels of data for this study. The first runs from 1993 to 1998, and the second runs from 1996 to 2001.

We select non-unionized workers between 20 and 64 years of age, working for positive wages, not in the public sector, and for whom we have observations for at least three consecutive years. We limit the sample to those who answered questions about wages, human capital variables, firm and establishment size. We dropped the information for the year 1993 in the first panel as the variable on wages was incomplete. These restrictions leave us with a sample of 21234 worker-year observations for which the first panel spans 5 years (1994-1998) and the second one 6 years (1996-2001).\textsuperscript{21}

\textsuperscript{20} We also add interactions between firm size and worker experience. We do so to proxy for the variance of $\theta_i$ which is an additional term entering the error term (5) when one takes the log of wages. Technically, in the full development of the statistical framework we use in the present paper, wages equal expected productivity which is an exponential function of $\theta_i$. If $\log \theta$ is normally distributed with mean $\mu$ and variance $\sigma^2$ then $E(\theta) = e^{\exp(\mu + 1/2\sigma^2)}$. See Gibbons, Katz, Lemieux and Parent (2005) for details on the derivation of the log wage equation.

\textsuperscript{21} Note that the two panels are independent in that there are no common individuals during the common period 1996-1998 of each panel. The panels are unbalanced because wage records may be missing for some years so that some individuals are observed for less than the 5 or 6 years of each panel.
The SLID provides information on two measures of employer’s size, namely the number of workers at all locations (firm size) and the number of workers at the respondent’s particular location (establishment size). Answers to both questions are categorized in the survey as follows: “Less than 20”, “20 to 99”, “100 to 499”, “500 to 999” and “1000 and over”. The theories of ability sorting and monitoring costs may involve both firm level and/or establishment level decision-making. We choose to concentrate on the analysis of firm size and estimate the returns to unmeasured ability by firm size as we think that decisions regarding ability sorting and monitoring are more likely to be taken at the firm level. Indeed, worker sorting and career opportunities are greater at the firm level compared to the establishment level. Also monitoring costs issues such as the trade-off between wages and supervision is part of the firm’s general staffing and compensation policies more likely to be consistently applied across establishments.

Furthermore, we find that the distribution of firm size using the SLID data is closer to that resulting from a Canadian employer based survey (the Survey of Employment, Payroll and Hours (SEPH)) than the distribution of establishment size. Morissette (1993) finds a similar result using the Labour Market Activity Survey. As a result, our analysis focuses on the relationship between firm size and wage outcomes but we replicate the final analysis using establishment size for the purpose of comparison. The comparison of firm size and establishment size effects provides an interesting check of our proposed conjecture: if wage policies in terms of the returns to skills and ability are determined at the firm level, the results for establishment size should be weaker than for firm size.

---

22 The specific questions are “How many persons were employed at the location where [you] work? and “About how many persons were employed at all these locations [where your employer operates]? With the same possible answers for each question namely: a) Less than 20 b) 20 to 99 c) 100 to 499 d) 500 to 999 e) 1000 and over f) Refuse to answer and g) Don’t know”.

23 Rebitzer (1995) finds evidence of the wage-supervision tradeoff estimating the impact of supervisory intensity on wages in the petrochemical industry where the analysis is done at the firm level.
Wage information is the hourly wage for a particular job at the end of the reference year. The amount includes tips, bonuses and commissions. For respondents who reported their wage or salary at this job as an hourly amount, the value is taken directly. For respondents who reported their wage or salary on some other basis, the amount is converted to an hourly “implicit” rate, based on all the relevant information available for that job/worker. Information regarding wages has been deflated to 1992 prices, using the Canadian CPI. In all the estimations we will use wages in log.

Experience is measured by a SLID derived variable that computes the full-time-full-year equivalent number of years of work experience. This definition includes all work (part time and full time) done since first starting to work full time. A value of zero is given for people with less than a year of experience and for those who never worked full-time.

Table 1 shows the different firm and worker characteristics by firm size. Consistent with findings from the literature on firm size and wages, workers are more experienced and more educated in larger firms. Wages are also higher in larger firms with a 48% pay differential between the smallest and the largest firm.

To provide a sense of the source of variation existent in the data, table 2a reports the transitional probabilities of firm size changes between two consecutive periods. The diagonal cells of the table correspond to the proportion of observations reflecting no changes in firm size between two consecutive years, the top right of the diagonal corresponds to changes to larger firms and the bottom left of the diagonal corresponds to changes to smaller firms. Comparing the diagonal cells to the total for each row of

---

24 When wages are reported using other units, a vector of the volume of hours paid for is created, taking into account start and end dates of the job, the number of usual hours worked/week, the number of months/year, the various work schedules, and unpaid absences.
25 CANSIM II series V735319.
26 It is therefore possible that we underestimate experience. We believe, however, that workers who never worked full time and with low attachment to the labour force merit a separate analysis.
size categories, one can see that a total of about 5-7% of the observations correspond to moves out of a given size category suggesting that worker mobility out of a given firm size category is relatively evenly distributed across the different firm size categories. Most, but not all changes occur across firms of consecutive size.

Table 2b shows the average hourly wage at time $t$ associated with firm size transitions between $t$ and $t - 1$ (wage in the new job for job changers). Comparing the cells on the diagonal with the cells to the right of the diagonal for each row size, it appears that movements to larger firms results in higher than average wages, while changes to smaller firms result in lower than average wages. This finding seems to be consistent with ability sorting into larger firms.

Changes in firm size between two periods may reflect the fact that the same firm is expanding or shrinking over time (due to substantial hiring or layoffs). We address this issue using information on job changes. We use the job identifier provided by the SLID to track down changes in all jobs held by a worker and compute a variable of job changes. Approximately 28% of all observations reporting a change in firm size corresponds to a change in job.

We also have information about whether the job changes are reported to be voluntary or involuntary. We further selected the sample of involuntary job changers and performed the estimations on this sample to minimize problems associated with

---

27 It may also be affected by mis-classifications in reported firm size between two survey years. We address this issue and explain it in details at the end of the results section.

28 The SLID includes a variable to indicate job changes (chjob28). Unfortunately, when several job changes occurred during a survey year, the most recent job change was often the only one recorded, resulting in an understatement of job changes.

29 The SLID questionnaire provides a question about reasons for the change in job and further classifies the reasons into voluntary and involuntary job changes. Involuntary job changes refer to the following reasons: 1-Company moved 2-Company went out of business 3-Seasonal nature of work-Layoff/Business slowdown (not caused by seasonal conditions) 3-Labour dispute 4-Dismissal by employer- 4-Temporary job/contract ended. Note that Employer dismissal may be related to ability and performance. We re-run the estimations without this category of workers and found similar results.
firm size changes due to firms changing size categories as a result of substantial hiring or layoffs. The use of involuntary job changers, or workers that changed job for exogenous reasons not related to ability or skills, also helps address sample selection biases affecting the estimates when voluntary job changers are included.  

3.2 The Size-Wage Gap

In this section we replicate the analysis performed in previous empirical studies of the links between firm size and wages with a comparison of the results when establishment size is used instead of firm size. Empirical studies on the determinants of wages and wage growth have analyzed possible departures from the basic Mincer-type wage equation that explains wages as a function of education and a quadratic in experience. Since the wage equation is originally derived from a model of optimal investment in human capital, it does not offer a complete representation of the labor market as it only describes the supply side of the market. Most studies add variables to the wage equation that describe and capture variations in wages related to the demand side of the labor market. In particular, industry, occupation, unionization, establishment and firm size dummies are the main variables employed. This type of analysis allows one to obtain an estimate of the average size-wage gap.

Panel A in Table 3 presents the results of OLS estimations of employer-size wage differentials using the SLID data, in the spirit of the empirical literature on firm size and wage outcomes. Overall, the results are similar to those reported elsewhere in the literature. The first column of coefficients shows the average size-wage gap in the absence of human capital variables.  

30 Note also that all our estimations are performed using the sample weights available for each panel. As a robustness check we compared the results with and without the weights and found no major differences.

31 These regressions include other standard demographic, labour market and geographical controls.
dummy for firm size reveal significant wage differentials (relative to firms with less than 20 employees) that range between 8% for firms with 20 to 99 employees to 25% for firms with more than a thousand employees. This is comparable to the estimates in Morissette (1993) using the Canadian Labour Market Activity Survey (LMAS) for 1986, which suggests that the magnitude of the gaps has not changed over time. The next column shows the results when controls for productive skills (education and a quadratic in experience and tenure) are added to the wage equation. The size of the coefficients slightly drops to 7% for firms with 20 to 99 employees and 20% for firms with more than a thousand employees. The third column shows the results of a fixed effect estimation of the same equation. The coefficients are substantially reduced but still positive and significant. In panel B we repeat the fixed-effect analysis using establishment size instead of firm size. Similar to the analysis based on firm size, the fixed-effect coefficients on establishment size are positive and still significant. As found in the literature based on U.S. data, the wage gap associated with larger establishments is in all cases greater than the firm size wage gap (Brown and Medoff, 1989).

In summary, the data used in this paper show very similar features to the data used in previous studies based on both Canadian and U.S. data. In the next section, we exploit the temporal dimension of our data set to investigate the potential importance of unmeasured ability. We estimate the statistical model presented in the previous section to emphasize the possibility of non random allocation of workers across firms of different sizes and estimate the differential returns to measured skills and unmeasured ability by firm size.

such as gender and race dummies as well as a dummy for whether the worker is married, industry and occupation dummies, years dummies and large city and region dummies.
4 Results

The most important finding in the analysis of the previous section is that the coefficients on firm size (and establishment size) from the fixed-effects estimations remain positive and statistically significant, suggesting that unmeasured/unobserved worker heterogeneity influences the wage gap but cannot explain all of it. The argument we use to understand the source of the remaining size-wage gap is based on the fact that if firms of different sizes have different wage policies for rewarding measured and unmeasured skills, the fixed-effects approach will not get rid of the influence of ability on wages. In the previous section we outlined two possible reasons why ability still affects the size-wage relationship depending upon what influences the wage determination process. The arguments have contrasting predictions about the existence of a size wage-premium. If the underlying mechanism for explaining the size-wage relationship comes from the fact that workers with a comparative advantage with respect to ability self-select into firms of different sizes, then the size-wage premium should disappear once ability is adequately controlled for. On the other hand, if the underlying mechanism is based on firms screening workers at the time of hire or monitoring on the job, then there should remain a premium as larger firms with expensive monitoring technology would offer greater wages (efficiency wages) to attract higher ability workers. Our analysis provides a way to identify ability effects (both measured and unmeasured) and the results can tell us whether a significant size-wage premium remains after such effects have been controlled for.

In this section we present the results of the estimation of wage equation (2) in which firm size is interacted with measured skills and unmeasured ability to account for different wage policies across firms of different sizes as a result of workers sorting based on comparative advantage and firm screening. The estimations are performed
over for the sample of non unionized private sector workers and the sub-sample of involuntary job changers. We are also interested in analyzing the size-wage profiles for the group of young workers for which sorting effects are likely to be stronger on wages. Therefore, we apply the estimations on the sample of involuntary job changers that are less than 35 years old to obtain and compare the estimates for the returns to measured skills and unmeasured ability with the estimates for the full sample of involuntary job changers.

In order to reduce the number of parameters to be estimated which multiplies with the interaction terms, we summarize workers measured skills into a skill index. To do so, we estimate a regression of the log wage on education, marital status, gender, race, a quadratic in experience and tenure, industry and occupation, union and public sector dummies, region, large city and year dummies for the entire original sample of workers. We then use the estimated coefficients related to education, and the quadratic in experience and tenure to compute the estimated or predicted log wage based on these human capital characteristics. The resulting skill index is normalized to 0.

The results of the GMM estimations are shown in table 4. The analysis is performed over the full sample of non unionized private sector workers (first column) and comparisons are made for different sub-samples. The results for involuntary job changers are shown in the second column. The results for the sample of involuntary job changers that are less than 35 years old are shown in the third column of table 4. Finally the results using establishment size instead of firm size are presented in the last column of the table for the sub-sample of involuntary job changers.
4.1 Returns to Unmeasured Ability and Measured Skills by Employer Size

The results for all non unionized private sector workers (first column of table 4) show significant evidence of non random sorting by firm size based on unmeasured ability. The test for equality of returns by firm size rejects the null at the ten percent level (p-value of 0.061). Interestingly, the returns are not monotonically increasing or decreasing with firm size as either the ability sorting or job screening would predict. Instead, the returns to ability in firms with 100-499 employees are significantly greater than one and therefore greater than the returns to unmeasured skills in firms with less than 20 employees\(^ {32}\) while the returns to unmeasured ability in firms larger than 500 employees are not significantly different from one. Looking at individual tests of equality of returns between consecutive sizes, we see that there is a clear rejection of the test of equality of returns between firms of size 100-499 compared to firms of size 20-99 suggesting that the returns to unmeasured ability are significantly greater in firms of size 100-499 compared to smaller firms. On the other hand, the returns are not significantly different from those in firms of size 500-999 or in firms of size 1000 or more employees. Overall, these first results suggests that comparative advantage based on unmeasured ability is important mainly for firms with 100-499 employees relative to smaller firms.

For measured skills, although the test of joint equality of the returns cannot reject the null, the individual test results show that the returns to measured skills are significantly lower in firms of size 100-499 relative to smaller firms and not different from the returns to measured skills in firms with 500 employees or larger\(^ {33}\). The

\(^{32}\)The returns to unmeasured ability in firms with less than 20 employees are set equal to one by construction.

\(^{33}\)Individual tests not shown but available upon request.
difference in the returns between firms of size 100-499 and larger firms is noticeable in magnitude and goes in the direction predicted by job screening (lower returns to ability and larger returns to measured skills in firms greater than 500 employees relative to smaller firms). The individual (pairwise) 2-sided tests of equality of the ratios of the returns cannot reject the null of equality for firms with 1000+ or 500+ employees compared to smaller firms (of size 20-99) at the 10% level (Wald statistic of values 2.52 and 2.13 respectively). On the other hand, the null is rejected against the one-sided alternative that the ratios are greater in firms of 500 to 999 employees compared to firms of size 100-499 (critical values of 1.83 and 2.32 for a 5% and a 1% level test respectively). In other words, the direction of the joint inequalities in the returns to both measured skills and ability is consistent with job screening ($\beta_{500-999} > \beta_{100-499}$ and $\lambda_{500-999} > \lambda_{100-499}$). These one-sided test results suggest the presence of greater job screening in larger firms (500 to 999 employees) relative to firms of less than 500 employees.

The next column, based on the sample of involuntary job changers, shows similar but more precise results. The non monotonicity in the returns to ability by firm size is more evident and there is again significant evidence of the importance of job screening in larger firms. Indeed, the returns to unmeasured ability in firms with 100-499 and 500-999 employees are significantly different from one. Moreover, the returns in firms with 500-999 are greater than those in both smaller and larger firms sizes (Wald statistics of values 7.32, 3.30 and 10.56 for differences with sizes 20-99, 100-499 and 1000+, respectively). In addition, the returns to unmeasured ability in the largest size category 1000+ are significantly smaller than in the category 100-499 (statistics value of 6.98). In terms of measured skills, the pairwise tests show that the equality of the ratios is rejected when comparing firms of size 500-999 to 20-99 and 1000+ to 20-99 (statistics values of 2.17 and 3.36 respectively) and the direction of
the inequalities is again consistent with job screening. Overall this suggests that while ability sorting based on comparative advantage is important for explaining workers sorting into medium-sized firms (500-999), job screening seems to be relatively more important for explaining workers sorting into firms of more than 500 employees. For these firms, it is possible that monitoring costs become too large and differentiating individual ability based on cash wages might be too expensive. These firms would therefore rely more on easily measurable skills (education, experience, tenure) to differentiate workers wage outcomes.

The third column corresponds to the same estimation for the sample of involuntary job changers aged 35 or less. In this case, although the overall test of equality in the returns to ability is rejected, the non-monotonicity still holds with respect to the returns to unmeasured ability, which are greater in firms of size 100-499 than in smaller firms and also greater than in firms with more than 500 employees. On the other hand, the inequality in the estimated ratios is not consistent with the restriction on the coefficients predicted by job screening. Overall, the results for younger job changers suggest that ability sorting is relatively more important in firms of size 100-499 employees and that there is no clear evidence of job screening effects. The weak evidence on job screening for younger job changers compared to all involuntary job changers may not be surprising. These workers are a more homogenous group in terms of work experience and so information on experience or tenure is not as informative as experience and tenure for the whole population of workers. Furthermore, the group of younger workers is smaller than the entire population of employees within a firm and therefore the size of the firm size and its impact on monitoring costs might be less of an issue for evaluating ability.

34This is likely to be due to the fact that except for the returns to 100-499 employees which are significantly greater than one, all other returns are close to one and not different from each other.
In terms of the estimated intercepts (bottom part of table 4), the $\alpha_j$s, and going across the column of the different sub-sample, the coefficients are insignificant in the estimations over the full sample of non unionized private sector workers which suggests that for such group of workers, previous firm-size wage premia found in the literature are indeed a result of non random ability sorting. The lack of significant intercepts is consistent with the fact that for that particular sample of workers, the returns to skills and unmeasured ability discussed previously show no clear evidence of job screening. For involuntary job changers including younger workers (columns 2 and 3), there is still evidence of a firm size wage premium (significant intercepts) in larger firms with 1000+ employees which is consistent with the previous results on the importance of job screening in firms of this size. The estimated coefficients suggest that job changers earn 10.9% more when joining a firm with 1000+ employees and earn 31.3% more when joining such firm when young.\(^{35}\) Surprisingly, there is also a size wage premium for firms with 100-499 employees. Job changers joining such firms seem to enjoy an efficiency wage in addition to greater rewards for both unmeasured ability and measured skills relative to job changers joining smaller and larger firms. Note that the time span for each panel is at most $5/6$ years which is a short period of analysis when considering job changers (especially younger workers). Although the estimated wage premium reflects an average wage differential over time between the wages in the old and in the new job, the short time period over which workers are followed and workers’s wages are analyzed may lead to the results that the intercept pick up potential signing bonuses at entry in the new job.

Panel B of the table shows the same estimations performed for the sample of involuntary job changers with establishment size instead of firm size. There is evidence of differential returns to ability by establishment size and also evidence of non mono-

\(^{35}\)As mentioned earlier, these intercept coefficients should be interpreted with caution as they are conditional on average ability which is unknown.
tonicity in these returns across size categories. In this case, the non monotonicity comes from the fact that the returns in firms of size 100-499 are not significantly different from one but the returns in firm of size 20-99 and 1000+ are significantly smaller than one. The tests of equality of the ratios do not reject the null which implies that we cannot draw conclusions on the presence of job screening. Interestingly, the returns to ability are systematically smaller than those estimated with firm size for the same sample of job changers (column 2 of the table). These results are consistent with our original conjecture that wage policies in terms of the returns to ability are more likely to be determined at the firm level, not the establishment level. Furthermore, it is possible that job changes occur across establishments of different sizes within the same firm. In this case, sorting and screening may have already occurred and wage changes would be smaller, and the wage structure by establishment size flatter, as a result.

In summary, the differential in the returns to both measured and unmeasured skills by firm size and the size-wage premium are sensitive to firm size. We find evidence of a non monotonic relationship between firm size and the returns to skills for both measured skills and unmeasured ability. The returns to unmeasured skills are significantly higher in medium-sized firms relative to smaller firms which is consistent with the ability sorting hypothesis. On the other hand, the returns to unmeasured skills are lower and the returns to measured skills generally higher in large firms relative to medium firms. This result for both measured skills and unmeasured ability is consistent with the screening hypothesis. Together these results suggest that there seems to be a firm size threshold after which monitoring costs are too high (or measurement of ability becomes too noisy) and screening based on measured skills is more advantageous. A similar result but at a theoretical level is conjectured in Long and Boylan (2003) who theorize that for a typically “S-shaped” production function
and given the tradeoff between specialization (in monitoring) objectives and agency issues increasing with firm size, the optimal level of monitoring should be higher in medium-size firms than in both larger and smaller firms.

We also find that ability sorting based on workers comparative advantage is important for younger workers switching to firms of size 100-499 employees and only weak evidence of screening based on monitoring and supervision for this group of workers. Finally we find that the implications of worker sorting in terms of comparative advantage and screening have a stronger effect at the firm level compared to the establishment level.

4.2 Robustness Checks

To assess the robustness of the preceding results, we tested the validity of the instruments used and the predictive power of the instruments. We also checked for the possibility of errors in the reported firm size categorization between two periods.

In terms of the validity of the instruments used to perform the estimation, we use the Hansen over-identification test and report the statistic and p-value at the bottom of table 4. In none of the cases can the validity of the instruments be rejected.

The analysis of the predictive power of the instruments is based on the computation of an F-test for the joint significance of the instruments when the instrumented variables (previous wages and firm size dummies) are regressed on the instruments, including all the exogenous variables in the right-hand side of the wage equation. The null hypothesis is therefore that the coefficients associated with the instruments are

\[\text{Hansen (1982) shows that } N \times \text{ minimized value of the objective function follows a } \chi^2 \text{ distribution with degrees of freedom equal to the number of over-identifying restrictions (number of instruments minus the number of parameters). In our case, it follows a } \chi^2 \text{ with 67 degrees of freedom.}\]
all identically equal to zero. The excluded instruments for the test are interactions in firm size affiliations between $t - 2$ and $t - 1$ as well as lagged versions of interactions between firm size affiliation and skills and experience. The results for the different samples are shown in appendix A table A1. One can conclude from the values of the F statistics for each sub-sample that the instruments have predictive power in explaining previous period wages and firm size choices.

The results might be contaminated by the presence of errors in firm size categorizations reported by individuals between two survey years. This type of mis-classification errors is likely to be serially uncorrelated. In this case, one way to check for the importance of such type of errors is to re-run the estimations with the wage equation computed in second quasi-difference, using the current period and the second lag (instead of the first lag) of the wages and right-hand side variables.\(^{37}\) Differences in the results would indicate that mis-classification errors might be severe. We were able to re-estimate the model with second differencing and found similar results in terms of the non monotonicity in the returns to both unmeasured ability and measured skills by firm size. We conclude from this that mis-classification errors in firm size categorization is not a main factor driving the variations we captured in our results.\(^{38}\)

\(^{37}\)This technique also involves using the third and second lags of the instruments.

\(^{38}\)Note that although we used the same estimation strategy, we encountered convergence problems due to the estimation of the control variables with little time variation (gender, race and marital status). To circumvent this problem, we redefined the skill index to reflect predicted wages based on gender, race and marital status in addition to the quadratic in experience and tenure. Doing so helped achieve convergence. We believe the different definition of measured skills does not alter the main interpretation of the model and the results. In fact, we re-ran the previous estimations in table 4 using this new skill index and found similar results as well. The results of the second quasi-difference estimations and of the original estimations based on the more individual-specific skill index are available upon request.
5 Conclusion

In this paper we re-examined the relationship between firm-size and wages by analyzing differences in the way firms of different sizes evaluate workers’ skills and in particular unmeasured ability. Using a GMM estimator, we are able to estimate the differential wage impact of unmeasured ability across firms of different sizes. Our data come from the Survey of Labour and Income Dynamics (SLID) over the years 1994 to 2001.

In terms of differences in wages, there is evidence that wages reflect the non random assignment of workers into firms of different sizes in accordance with the ability sorting argument. At the same time, there are important differences in the wage policies of firms of 1000 employees or more compared to medium-size firms which suggests the importance of monitoring and screening issues.

The results of the GMM estimations show that there seems to be a non monotonic relationship between firm size and the returns to skills for both measured and unmeasured skills. While the returns to unmeasured ability are greater in firms with size 100-499 compared to smaller firms, in line with the ability sorting hypothesis, they are also found to be greater than in firms with 1000 employees or more. In addition, the returns to measured skills are significantly greater in these larger firms compared to medium size firms (with 100-499 employees) suggesting that firms with more than 1000 employees may face monitoring and screening issues. This result on both measured and unmeasured skills is consistent with the screening hypothesis.

The results are stronger in terms of significance and magnitude of the coefficients for the sample of involuntary job changers (for which firm and establishment size changes are the result of true job changes). We view this as a confirmation of the robustness of the results. For younger workers (and job changers), the evidence of
differential wage policies by firm size is mainly explained by ability sorting into larger firms based on comparative advantage with only weak evidence of firm screening. Firms with 100-499 employees seem to reward both unmeasured ability and measured skills more than smaller firms. Interestingly, for these firms there is still a wage premium suggesting that they also pay efficiency wages to attract higher ability workers. It is possible that for these young job changers observed over the 5/6 year period of our sample, the premium reflect signing bonuses.

From a worker’s standpoint, the results suggest that the decision to join a small, medium or large firm seems to depend on the relative level of measured versus unmeasured skills. For example higher ability workers with also greater education and/or experience have a greater advantage (in terms of wages) in joining a large firm while higher ability workers with less education and/or experience have a comparative advantage in medium-size firms compared to larger or smaller firms.

The results of this paper are based on survey data and it would be interesting to replicate the analysis using matched employer-employee data in order to be able to isolate firm-size effects from other firm-specific effects. Given the result on the non monotonic relationship in the returns to measured and unmeasured ability by firm size, an area for future work would be to use firm size as a continuous variable (whenever available on a longitudinal basis) and obtain a more precise estimate of the firm size threshold for better understanding firm decisions concerning screening, monitoring and evaluation of ability. Finally, an additional extension of this work would be to study worker sorting in relation to differences in compensation policies by firm size, including for example information on variable pay (stock options, profit sharing) and on fringe benefits.
References


Figure 1: Optimal Worker Sorting in the Comparative Advantage Model

Figure 2: Optimal Worker Sorting in the Screening Model
Table 1. Average Sample Characteristics
Non Unionized Workers in the Private Sector

<table>
<thead>
<tr>
<th>Workforce Characteristics (%)</th>
<th>All</th>
<th>0-19</th>
<th>20-99</th>
<th>100-499</th>
<th>500-999</th>
<th>1000+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less Than High School</td>
<td>12.7</td>
<td>17.2</td>
<td>16.5</td>
<td>11.2</td>
<td>9.2</td>
<td>7.4</td>
</tr>
<tr>
<td>High School</td>
<td>30.3</td>
<td>32.5</td>
<td>27.8</td>
<td>28.0</td>
<td>29.6</td>
<td>31.3</td>
</tr>
<tr>
<td>Post Secondary</td>
<td>57.0</td>
<td>50.3</td>
<td>55.7</td>
<td>60.8</td>
<td>61.2</td>
<td>62.0</td>
</tr>
<tr>
<td>Female Dummy</td>
<td>45.9</td>
<td>53.7</td>
<td>40.9</td>
<td>41.2</td>
<td>43.6</td>
<td>44.9</td>
</tr>
<tr>
<td>Experience (Yrs)</td>
<td>16.1</td>
<td>15.3</td>
<td>16.3</td>
<td>16.6</td>
<td>16.7</td>
<td>16.6</td>
</tr>
<tr>
<td>Tenure (Yrs)</td>
<td>9.3</td>
<td>8.1</td>
<td>8.5</td>
<td>9.0</td>
<td>9.8</td>
<td>11.2</td>
</tr>
<tr>
<td>Hourly Wage</td>
<td>14.46</td>
<td>11.60</td>
<td>13.96</td>
<td>15.57</td>
<td>16.03</td>
<td>17.13</td>
</tr>
<tr>
<td>Observations</td>
<td>21,275</td>
<td>6,719</td>
<td>4,246</td>
<td>2,802</td>
<td>1,481</td>
<td>5,986</td>
</tr>
</tbody>
</table>
### Table 2a Transition Probabilities (%)

<table>
<thead>
<tr>
<th>Firm Size (t-1)</th>
<th>0-19</th>
<th>20-99</th>
<th>100-499</th>
<th>500-999</th>
<th>1000+</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-19</td>
<td>25.4</td>
<td>3.6</td>
<td>1.0</td>
<td>0.4</td>
<td>1.0</td>
<td>31.3</td>
</tr>
<tr>
<td>20-99</td>
<td>3.7</td>
<td>12.1</td>
<td>2.4</td>
<td>0.6</td>
<td>1.3</td>
<td>19.7</td>
</tr>
<tr>
<td>100-499</td>
<td>1.2</td>
<td>2.3</td>
<td>6.7</td>
<td>1.4</td>
<td>1.9</td>
<td>13.4</td>
</tr>
<tr>
<td>500-999</td>
<td>0.5</td>
<td>0.6</td>
<td>1.4</td>
<td>2.2</td>
<td>2.6</td>
<td>7.3</td>
</tr>
<tr>
<td>1000+</td>
<td>1.3</td>
<td>1.4</td>
<td>1.7</td>
<td>2.4</td>
<td>21.4</td>
<td>28.2</td>
</tr>
<tr>
<td>TOTAL</td>
<td>31.6</td>
<td>20.0</td>
<td>13.2</td>
<td>7.0</td>
<td>28.2</td>
<td>100.0</td>
</tr>
</tbody>
</table>

### Table 2b Average Wage by Transitions

<table>
<thead>
<tr>
<th>Firm Size (t-1)</th>
<th>0-19</th>
<th>20-99</th>
<th>100-499</th>
<th>500-999</th>
<th>1000+</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-19</td>
<td>11.51</td>
<td>12.50</td>
<td>12.59</td>
<td>11.88</td>
<td>12.27</td>
<td>11.68</td>
</tr>
<tr>
<td>100-499</td>
<td>11.83</td>
<td>14.66</td>
<td>16.15</td>
<td>15.86</td>
<td>16.73</td>
<td>15.58</td>
</tr>
<tr>
<td>500-999</td>
<td>11.23</td>
<td>13.73</td>
<td>15.04</td>
<td>16.87</td>
<td>16.58</td>
<td>15.77</td>
</tr>
<tr>
<td>1000+</td>
<td>11.42</td>
<td>14.18</td>
<td>16.31</td>
<td>16.18</td>
<td>17.60</td>
<td>16.94</td>
</tr>
<tr>
<td>TOTAL</td>
<td>11.60</td>
<td>13.96</td>
<td>15.57</td>
<td>16.03</td>
<td>17.13</td>
<td>14.46</td>
</tr>
</tbody>
</table>
### Table 3. Estimates of the Size Wage Gap
Non Unionized Workers in the Private Sector

<table>
<thead>
<tr>
<th></th>
<th>Panel A. Firm Size</th>
<th></th>
<th></th>
<th>Panel B. Establishment Size</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
<td>OLS</td>
<td>Fixed-Effects</td>
<td>OLS</td>
<td>Fixed-Effects</td>
</tr>
<tr>
<td>Firm Size 20-99</td>
<td>0.081***</td>
<td>0.070***</td>
<td>0.011*</td>
<td>0.085***</td>
<td>0.017***</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.010)</td>
<td>(0.007)</td>
<td>(0.009)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Firm Size 100-499</td>
<td>0.163***</td>
<td>0.136***</td>
<td>0.027***</td>
<td>0.188***</td>
<td>0.035***</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.012)</td>
<td>(0.008)</td>
<td>(0.010)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Firm Size 500-999</td>
<td>0.202***</td>
<td>0.167***</td>
<td>0.037***</td>
<td>0.243***</td>
<td>0.045***</td>
</tr>
<tr>
<td></td>
<td>(0.016)</td>
<td>(0.015)</td>
<td>(0.009)</td>
<td>(0.023)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Firm Size 1000+</td>
<td>0.226***</td>
<td>0.181***</td>
<td>0.045***</td>
<td>0.347***</td>
<td>0.061***</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.010)</td>
<td>(0.008)</td>
<td>(0.021)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>High School</td>
<td>--</td>
<td>0.067***</td>
<td>-0.026</td>
<td>0.071***</td>
<td>-0.025</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.011)</td>
<td>(0.032)</td>
<td>(0.010)</td>
<td>(0.032)</td>
</tr>
<tr>
<td>Post Secondary</td>
<td>--</td>
<td>0.233***</td>
<td>0.001</td>
<td>0.227***</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.010)</td>
<td>(0.033)</td>
<td>(0.010)</td>
<td>(0.033)</td>
</tr>
<tr>
<td>Experience</td>
<td>--</td>
<td>0.021***</td>
<td>0.050***</td>
<td>0.021***</td>
<td>0.051***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.002)</td>
<td>(0.006)</td>
<td>(0.002)</td>
<td>(0.006)</td>
</tr>
<tr>
<td>Exp2 (/100)</td>
<td>--</td>
<td>-0.038***</td>
<td>-0.062***</td>
<td>-0.039***</td>
<td>-0.062***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.004)</td>
<td>(0.007)</td>
<td>(0.004)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>Tenure</td>
<td>--</td>
<td>0.017***</td>
<td>0.012***</td>
<td>0.017***</td>
<td>0.012***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Ten2 (/100)</td>
<td>--</td>
<td>-0.026***</td>
<td>-0.024**</td>
<td>-0.025***</td>
<td>-0.023**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.006)</td>
<td>(0.011)</td>
<td>(0.004)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.36</td>
<td>0.45</td>
<td>0.17</td>
<td>0.46</td>
<td>0.17</td>
</tr>
</tbody>
</table>

**Note:** Additional controls include indicators for gender, race, married, a dummy for living in a large city, geographical area of residence, occupation, industry and survey year. The number of observations is 21234.

*** Indicates the coefficient is significant at 1%, ** indicates that the coefficient is significant at 5%, * indicates that the coefficient is significant at 10%
Table 4. Returns to Unmeasured Skills by Employer Size  
Non Unionized Workers in the Private Sector

<table>
<thead>
<tr>
<th></th>
<th>Panel A. Firm Size</th>
<th></th>
<th>Panel B. Establishment Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All Workers</td>
<td>Involuntary Job Changers</td>
<td>Involuntary Job Changers</td>
</tr>
<tr>
<td>Returns to Unmeasured Skills ( \lambda_j )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size 20-99</td>
<td>0.899***</td>
<td>0.974***</td>
<td>0.846***</td>
</tr>
<tr>
<td></td>
<td>(0.194)</td>
<td>(0.186)</td>
<td>(0.154)</td>
</tr>
<tr>
<td>Size 100-499</td>
<td>1.671***†</td>
<td>1.667***†</td>
<td>1.515***†</td>
</tr>
<tr>
<td></td>
<td>(0.388)</td>
<td>(0.394)</td>
<td>(0.305)</td>
</tr>
<tr>
<td>Size 500-999</td>
<td>1.473***</td>
<td>2.613***†</td>
<td>0.979***</td>
</tr>
<tr>
<td></td>
<td>(0.424)</td>
<td>(0.659)</td>
<td>(0.299)</td>
</tr>
<tr>
<td>Size 1000 +</td>
<td>1.359***</td>
<td>0.789***</td>
<td>0.972***</td>
</tr>
<tr>
<td></td>
<td>(0.336)</td>
<td>(0.204)</td>
<td>(0.150)</td>
</tr>
</tbody>
</table>

Tests for Equality of Returns \( \lambda_{100-499} = \lambda_{20-99}, \ldots, \lambda_{1000+} = \lambda_{500-999} \)  
(p-value)  

|                      |                         |                          |                            |
|                      | 8.98                  | 20.35                    | 6.08                       | 12.64                     |
|                      | (0.061)               | (0.000)                  | (0.193)                    | (0.013)                   |

Returns to Measured Skills \( \beta_j \)  
Size <20  
Size 20-99  
Size 100-499  
Size 500-999  
Size 1000 +  

Tests for Equality of Returns \( \beta_{100-499} = \beta_{20-99}, \ldots, \beta_{1000+} = \beta_{500-999} \)  
(p-value)  

|                      |                         |                          |                            |
|                      | 5.36                  | 8.21                     | 10.14                      | 7.34                       |
|                      | (0.251)               | (0.08)                   | (0.038)                    | (0.11)                     |

NOTE: All specifications include controls for gender, race, marital status, industry, occupation, city, region and year. The tests for equality of returns are \( \chi^2 \) tests. ***: < .01, **: < .05, *: < .10. The symbol † means that the coefficient is significantly different from one. The full sample contains 21234 observations. The sample of involuntary job changers has 1750 observations. The sample of young workers job changers has 697 observations.
Table 4. Returns to Unmeasured Skills by Employer Size
Non Unionized Workers in the Private Sector
- Continued -

<table>
<thead>
<tr>
<th>Test for Equality of Ratios</th>
<th>Panel A. Firm Size</th>
<th>Panel B. Establishment Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All Workers</td>
<td>Involuntary Job Changers</td>
</tr>
<tr>
<td>( \beta_{100-499}/\lambda_{100-499} = \beta_{20-99}/\lambda_{20-99} )</td>
<td>6.48</td>
<td>0.55</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.01)</td>
<td>(0.456)</td>
</tr>
<tr>
<td>( \beta_{500-999}/\lambda_{500-999} = \beta_{20-99}/\lambda_{20-99} )</td>
<td>2.13</td>
<td>0.01</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.144)</td>
<td>(0.923)</td>
</tr>
<tr>
<td>( \beta_{500-999}/\lambda_{500-999} = \beta_{100-499}/\lambda_{100-499} )</td>
<td>0.44</td>
<td>0.95</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.507)</td>
<td>(0.328)</td>
</tr>
<tr>
<td>( \beta_{1000+}/\lambda_{1000+} = \beta_{20-99}/\lambda_{20-99} )</td>
<td>2.52</td>
<td>2.17</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.112)</td>
<td>(0.09)</td>
</tr>
<tr>
<td>( \beta_{1000+}/\lambda_{1000+} = \beta_{100-499}/\lambda_{100-499} )</td>
<td>0.48</td>
<td>2.23</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.487)</td>
<td>(0.135)</td>
</tr>
<tr>
<td>( \beta_{1000+}/\lambda_{1000+} = \beta_{500-999}/\lambda_{500-999} )</td>
<td>0.00</td>
<td>3.36</td>
</tr>
<tr>
<td>(p-value)</td>
<td>(0.960)</td>
<td>(0.066)</td>
</tr>
</tbody>
</table>

**Intercept \( \alpha_i \)**

| Size 20-99 | 0.013 | 0.028 | 0.010 | -0.011 |
| Size 100-499 | 0.010 | 0.027 | 0.418** | 0.004 |
| Size 500-999 | 0.023 | 0.426 | 0.197 | 0.027 |
| Size 1000+ | 0.026 | 0.109* | 0.313*** | 0.010 |

**Overidentification Test**

| (p-value) | 81.25 | 49.89 | 41.33 | 53.40 |

**NOTE:** All specifications include controls for gender, race, marital status, industry, occupation, city, region and year. The tests for equality of returns are \( \chi^2 \) tests. ***: < .01, **: < .05, *: < .10. The symbol † means that the coefficient is significantly different from one. The full sample contains 21234 observations. The sample of involuntary job changers has 1750 observations. The sample of young workers job changers has 697 observations.
Appendix A: Instruments

The statistical framework we use enables us to estimate differential returns to unmeasured ability and measured skills by firm size. In addition to treating the issues of non random selection and quasi-differencing (and resulting instrumentation of lagged wages), the framework and estimation method also deals with the issue of endogenous mobility. In this case, current choice of firm size needs to be instrumented in addition to lagged wages.

The intuition for the framework that includes endogenous mobility and the resulting choice of instruments goes the following way. Endogenous firm size switching occurs as a result of individuals learning about their own ability and how suited it is to current firm size relative to other firm size opportunities (which may reward more or less ability). Expected ability is updated every period based on the observation of the worker’s previous period output. Future expected ability increases (decreases) as a result of a positive (negative) productivity shock in the current period output which could not have been anticipated previously (rational expectations). If the explanation of pay settings based on workers comparative advantage prevails (over the job screening hypothesis), a large enough increase in expected ability leads workers to switch to a larger firm.

In this setting, expected ability follows a martingale process (rational expectations). The martingale assumption implies that changes in expected ability are not serially correlated. Using this assumption and the exogeneity condition (8), the history of previous period firm size assignments provides a valid candidate for instrumenting current choice of firm size as it is uncorrelated with the error term in the wage equation (which in this case includes the error term of the martingale process for expected ability). It is also a good candidate for helping predict current choice of firm size as it gives information about previous period expected ability which is used to predict current choice of firm size.

Table A1 below shows the results of tests of the predictive power of the instruments for the sample of non unionized private sector workers. The number of excluded instruments is 40.

<table>
<thead>
<tr>
<th></th>
<th>Lag Wage</th>
<th>Size 20-99</th>
<th>Size 100-499</th>
<th>Size 500-999</th>
<th>Size 1000 +</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>F stat</strong></td>
<td>11.82</td>
<td>10.23</td>
<td>11.09</td>
<td>7.41</td>
<td>9.12</td>
</tr>
<tr>
<td><strong>(p-value)</strong></td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
</tbody>
</table>