INTRA-FIRM WAGE BARGAINING AND EMPLOYEE HOLD-UP:

A TEST OF THE OVER-EMPLOYMENT HYPOTHESIS

Yongseung Han
Gainesville State College

Arthur Snow
University of Georgia

Ronald S. Warren, Jr.*
University of Georgia

Abstract

We develop and implement tests of two predictions of the Stole-Zwiebel model of intra-firm bargaining in the presence of employee hold-up, using panel data on savings banks in Korea. We first test for the presence of allocative inefficiency, and find evidence of the over-employment of labor. This result is consistent with a key prediction of the Stole-Zwiebel model, and runs counter to the competing under-employment hypothesis offered by de Fontenay and Gans. We then estimate the "front-load" factors for the firms in our sample, and cannot reject the Stole-Zeiebel prediction that all of these front-load factors are unity. Our empirical results thus provide evidence in support of the theory that firms respond to the prospect of employee hold-up by hiring additional workers until the bargaining-determined wage equals the workers' outside option, thereby eliminating the quasi-rents accruing to incumbent employees with specialized knowledge.

JEL Classifications: D20 · L20
*We thank Harold Mulherin for helpful comments on an earlier draft.

INTRA-FIRM WAGE BARGAINING AND EMPLOYEE HOLD-UP:

A TEST OF THE OVER-EMPLOYMENT HYPOTHESIS

1. Introduction

The benchmark model for analyzing non-competitive wage setting within the firm is provided by Stole and Zwiebel (1996a, 1996b, 2003), who emphasize the importance of the bargaining power of incumbent workers who cannot readily be replaced by newly hired outsiders. They argue that, where current employees possess hold-up power arising from the specificity of their human capital, firms respond by hiring additional workers to dilute their individual bargaining power, leading to allocative inefficiency in the form of over-employment. Specifically, profit maximization in this environment dictates that workers are hired until the bargaining-determined wage equals the workers' outside option (i.e., the competitive wage), so that the quasi-rents accruing to the specific human capital vanish. As a by-product of this allocative inefficiency, technical inefficiency may also arise to offset firms' reliance on over-employment as the sole means of reducing the bargaining power of workers.

The Stole-Zwiebel model has been used to motivate a wide variety of theoretical analyses of wage determination in a bargaining environment. Cahuc and Wasmer (2001) incorporate the Stole-Zwiebel framework into a standard matching model of unemployment, and establish conditions under which workers are paid more than their reservation wage. Bertola and Garabaldi (2001) use the Stole-Zwiebel model to explain how intra-firm bargaining can lead large firms to pay high wages. Rebitzer and Taylor (2007) motivate their study of the organizational design of law firms by appealing to the

Stole-Zwiebel model and broaden the interpretation of specific human capital to encompass "knowledge assets" as a factor determining the outcome of intra-firm bargaining. More recently, Chakrabarti and Tangsangasaksri (2011) extend the Stole-Zwiebel model by allowing the firm's employees to organize in a finite number of unions and, in this context, examine the properties of stable coalitions under sequential, bilateral bargaining.

De Fontenay and Gans (2003) modify the bargaining environment in the Stole-Zwiebel model by introducing the possibility of replacing incumbent workers ("insiders") with "outsiders" to dilute the insiders' hold-up power. Thus, while Stole and Zwiebel assume that incumbent workers cannot be replaced by new hires from outside the firm, de Fontenay and Gans assume that incumbents and outsiders are perfect substitutes. Since replacing insiders with outsiders requires negotiated wages that are higher than the marginal product of labor at the neoclassical level of employment, de Fontenay and Gans (2003) predict that the wage-bargaining firm under-employs workers, in direct contradiction to the Stole-Zwiebel hypothesis. We develop an empirical strategy that allows us to discriminate between the Stole-Zwiebel over-employment hypothesis and the opposing under-employment hypothesis of de Fontenay and Gans.

We implement our tests using panel data on Korean savings banks. Table 1 presents descriptive statistics for the savings banks in our sample, along with comparable summary statistics for Korean commercial banks. Relative to commercial banks, savings banks in Korea employ more workers per branch and pay their workers considerably less, on average, but earn lower profits per worker. Of course, the typical Korean commercial bank is much larger than the average savings bank, and therefore has many more branches and many more workers in total. The average annual profitability of savings banks, as measured by the rate of return on assets (ROA) is, with the exception of 2006, lower than that of commercial banks.

The Stole-Zwiebel hypothesis may explain why savings banks in Korea hire more workers per branch than do commercial banks, even though the average worker at a savings bank contributes less to profit than his or her counterpart at a commercial bank. Loan officers at savings banks have a comparative advantage in "relationship banking" because of the personal knowledge they have concerning the credit-worthiness of the small businesses that comprise the majority of their potential borrowers. These small businesses typically do not have transparent accounting systems in place, so accurately evaluating their income statements and balance sheets requires "local" information that savings banks are able to acquire and process more effectively than their commercial-bank competitors. This specialized information resides in the savings banks' loan officers and is a valuable asset to both them and their employers. Once employed, however, loan officers at savings banks are in a position to expropriate the quasi-rents arising from the customer-specific information they possess by bargaining for a higher wage or threatening to quit and take their "knowledge capital" with them. According to the Stole-Zwiebel model, savings banks are predicted to anticipate the potential hold-up power possessed by these specialized employees and hire more of them than is dictated by neoclassical profit maximization. As a result, savings banks are predicted to hire "too many" workers, seemingly sacrificing profits in the process. By contrast, the de Fontenay-Gans insider-outsider model predicts "too few" workers are hired.

Our application of the Stole-Zwiebel model to a situation where mobile "knowledge workers," once employed, are in a position to hold up the firm is similar to that of Rebitzer and Taylor (2007), who examine the consequences of client-specific knowledge for the organizational structure of law firms. They argue that experienced lawyers have detailed, valuable information about a law firm's clients and are therefore in a position to appropriate a share of the firm's profits by threatening to "grab and leave" with these clients. Because law firms cannot readily establish property rights over this specialized knowledge, they attempt to mitigate the threat of hold up by conducting an up-or-out tournament among junior associates and then granting partner status to the winners which conveys an equity stake that becomes worthless if the partner subsequently leaves the firm. Thus, the firm in their model uses the terms of the employment contract, rather than the number of workers hired, to mitigate the hold-up problem.

Our test of the Stole-Zeiebel hypothesis proceeds in two steps. We first adopt a shadow-price approach and impose the assumption that technology is characterized by decreasing returns to scale in the variable inputs. This framework allows us to test for the presence of allocative inefficiency and, in particular, to test a key prediction of the Stole-Zwiebel model, namely that firms subject to employee hold-up over-employ labor.

It is important to note that a finding of over-employment is sufficient to reject the de Fontenay and Gans (2003) model. However, it constitutes only a necessary condition for the Stole-Zwiebel hypothesis to be consistent with the data. The distortions introduced in the Stole-Zwiebel model by profit-maximizing responses to potential hold up are characterized by a single statistic they call the front-load factor, which equals the profit of a Stole-Zwiebel firm relative to the profit of its neoclassical counterpart. Firms are predicted to increase employment until the wage is driven down to the workers' reservation wage, where the front-load factor is equal to one. Accordingly, if there is evidence of over-employment, we must then proceed to a second step and estimate the value of each firm's front-load factor.

To preview our empirical results, we first provide evidence of allocative inefficiency in the form of over-employment that is counter to the prediction of the de Fontenay-Gans model, but is in line with one of the predictions of the Stole-Zwiebel model. We then turn to the estimation of each firm's front-load factor. We cannot reject the null hypothesis that every firm's front-load factor is equal to one. Accordingly, we conclude that the firms in our sample behave in a manner consistent with the Stole-Zwiebel hypothesis by hiring excess labor until the bargained wage is equal to the competitive wage.

In the next section, we formally set out the key predictions of the Stole-Zwiebel model. In section 3, we provide the details of our strategy for testing these predictions. In section 4, we describe our data and discuss our estimation procedure. The empirical results are presented in section 5, and conclusions are offered in section 6.

2. The Stole-Zwiebel Hypothesis

Stole and Zwiebel (1996a) view the intra-firm bargaining process as essential to understanding a firm's hiring decisions and choice of technology. Intra-firm bargaining in their model determines employees' wages and the firm's profit, and takes place under the assumption that workers and the firm split the joint surplus, defined as revenue net of non-labor costs relative to their respective outside options. The outside option for an employee is the reservation wage, or the competitively determined market wage, $\underline{w}$, and the outside option for the firm is the outcome of a bargaining process with one less employee in the firm. In an environment with n identical employees, equal bargaining power between the firm and its employees implies that

$$\tilde{\pi}(n)-\tilde{\pi}(n-1)=\tilde{w}(n)-\underline{w}, \tag{1}$$

where $\tilde{\pi}(n)=F(n)-\tilde{w}(n)n$ is the firm's profit given the output (revenue) *F(n)*, and $\tilde{w}(n)$ denotes an employee's wage. Equilibrium wage and profit are then

$$\tilde{w}(n)=\frac{1}{n(n+1)}\sum_{i=0}^{n} iF_i(i)+\frac{1}{2}\underline{w} \tag{2a}$$

and

$$\tilde{\pi}(n)=\frac{1}{(n+1)}\sum_{i=0}^{n}\pi(i), \tag{2b}$$

where $F_n(n)$ denotes the marginal product of labor, and $\pi(n)=F(n)-\underline{w}n$ denotes neoclassical profit with n employees.

The firm maximizes its payoff, given by (2b), which is the uniform average of neoclassical profit as employment varies over $i=0,\cdots,n$ workers. The firm chooses levels of labor n and capital x to solve the problem

$$\max_{n,x}\tilde{\pi}(n,x)=\frac{1}{(n+1)}\sum_{i=1}^{n}\pi(i,x). \tag{3}$$

Stole and Zwiebel conclude that the wage-bargaining firm acts as a neoclassical firm with the induced production function

$$\tilde{F}(n,x)=\frac{1}{n+1}\sum_{i=0}^{n}F(i,x)+\frac{1}{2}\underline{w}n, \tag{4}$$

which includes the market wage $\underline{w}$ as a parameter.

The first-order conditions for the optimal levels of inputs $(\tilde{n}^*, \tilde{x}^*)$ are given by

$$\pi(\tilde{n}^*, \tilde{x}^*) = \tilde{\pi}(\tilde{n}^*, \tilde{x}^*) \tag{5a}$$

$$\sum_{i=0}^{\tilde{n}^*} \frac{\partial \pi(i, \tilde{x}^*)}{\partial x} = 0, \tag{5b}$$

indicating that the wage-bargaining firm employs labor and capital (or other inputs) up to the point where its profit is equal to the profit that would be earned by a neoclassical firm employing the same quantities of labor and other inputs, and the average of the marginal returns to other inputs is driven to zero. The equality of profits at $\tilde{n}^*$ expressed in (5a) implies that the wage paid by wage-bargaining and neoclassical firms must be equal as well. That is, the Stole-Zwiebel firm hires labor until the bargaining-determined wage is driven down to the market-determined wage, at which point employees lose the quasi-rents that would have been gained by *ex post* opportunistic bargaining. Thus, the profit-maximizing response of the firm in anticipation of potential employee hold up causes the wage premium to vanish and results in the over-employment of labor relative to the neoclassical profit-maximizing outcome, as illustrated in Figure 1. One implication of (5b) is that other inputs such as capital are under-utilized.

The input distortions introduced through the intra-firm bargaining process can be characterized by a single statistic γ that Stole and Zwiebel call the front-load factor, defined as

$$0 \le \gamma(F, n) \equiv 1 - \frac{1}{\pi(n)} \sum_{i=0}^{n} \frac{i}{n+1} \Delta\pi(i) \le 1, \tag{6a}$$

where Δ is the first-difference operator. The front-load factor measures the extent to which neoclassical profit margins are realized earlier in the production process. Stole and Zwiebel show that

$$\gamma(F,n) = \frac{\tilde{\pi}(n)}{\pi(n)}. \tag{6b}$$

Thus, at a given level of employment, (6b) indicates that firms prefer technologies with higher front-loading[1]. It follows that firms may choose an inefficient technology with a higher front-load factor rather than an efficient technology with a lower front-load factor. Finally, (5a) and (6b) together imply that the Stole-Zwiebel firm achieves maximum profit when the front-load factor is equal to one. We test this prediction of the model in the second step of our empirical examination of the Stole-Zwiebel hypothesis.

3. Testing the Stole-Zwiebel Hypothesis

The induced production function (4) implied by the Stole-Zwiebel hypothesis does not inherit all of the standard duality properties because of the presence of the market wage $\underline{w}$. For example, the minimum cost function associated with (4) is not homogenous of degree one in input prices. Therefore, the predictions of allocative and technical inefficiencies cannot be tested using conventional duality-based techniques for estimating cost and production functions. We overcome this difficulty by adopting a shadow-price approach and imposing an assumption of decreasing returns to scale in the variable inputs. To test the predictions of the Stole-Zwiebel model, we develop and implement a feasible, two-step procedure. The first step determines whether there is

[1] In terms of Figure 1, higher front-loading would lead to a steeper rise in the Stole-Zwiebel profit curve, with its peak occurring at an intersection higher on the neoclassical profit curve.

allocative inefficiency in the use of variable inputs with over-employment of labor. Given a finding of allocative inefficiency in the first step, the second step determines whether all firms' front-load factors equal one.

A Test for Allocative Inefficiency

Our test for allocative inefficiency involves a comparison of two models built on the following assumptions. In the restricted model, either neoclassical profit maximization holds or it fails to hold due solely to technical inefficiency. In the unrestricted model, neoclassical profit maximization may fail to hold due to allocative inefficiency or to both allocative and technical inefficiency. If the restricted model is correct, then there is no allocative inefficiency and the wage-determination process generating the data does not distort input choices. If the unrestricted model is correct, then there is allocative inefficiency. A log-likelihood-ratio test of the parameter restrictions implied by the Stole-Zwiebel hypothesis is performed to determine whether allowing for allocative inefficiency better explains the data.

The Restricted Model

In the restricted model, the firm's profit function is specified as

$$\pi(p,w,u) \equiv \max_{v} \left\{ py - w \cdot v;\ ye^{-u} = F(v) \right\}, \tag{7a}$$

where y and p are output quantity and per-unit output price, respectively, v and w are input quantities and their respective prices, and F denotes the technology. Following Kumbhakar (2001), we allow for the possibility that the firm is technically inefficient by introducing the factor e^{-u}. If $u = 0$, then $e^{-u} = 1$, $y = F(v)$, and the firm is technically efficient; if $u < 0$, then $e^{-u} > 1$, $y < F(v)$, and the firm is technically inefficient, operating in the interior of its technology set. With this addition, the profit function can be rewritten as

$$\pi(pe^{u},w)\equiv \underset{v}{Max}\left\{\left(pe^{u}\right)y-w\cdot v;\ y=F(v)\right\}. \tag{7b}$$

This is the firm's actual profit function, which we approximate with the translog form

$$\begin{aligned}\ln\pi(pe^{u},w)&=\alpha_0+\sum_{k=1}^{K}\alpha_k\ln w_k+\alpha_y\ln\left(pe^{u}\right)\\&+\frac{1}{2}\left[\sum_{k=1}^{K}\sum_{j=1}^{K}\alpha_{kj}\ln w_k\ln w_j+\alpha_{yy}\ln\left(pe^{u}\right)\ln\left(pe^{u}\right)\right]\\&+\sum_{k=1}^{K}\alpha_{ky}\ln w_k\ln\left(pe^{u}\right),\end{aligned} \tag{8a}$$

where the symmetry restrictions $\alpha_{kj}=\alpha_{jk}$ and $\alpha_{ky}=\alpha_{yk}$ are imposed. A random noise component, denoted by ε, is added to (8a) in estimation. By applying the Shephard-Uzawa-McFadden lemma to (8a), we obtain the cost-share equations

$$S_k=-\alpha_k-\sum_{j=1}^{K}\alpha_{kj}\ln w_j-\alpha_{ky}\ln\left(pe^{u}\right)\qquad \forall\ k=1,\cdots,K \tag{8b}$$

with a random noise component ε_k added in estimation. To avoid singularity, we omit the revenue-share equation and estimate the system of $K+1$ equations given in (8a) and (8b). Since the profit function is homogenous of degree one in input prices and the technical-inefficiency-adjusted output price (pe^{u}), we impose the parameter restrictions

$$\sum_{k=1}^{K}\alpha_k+\alpha_y=1,\ \sum_{k=1}^{K}\alpha_{ky}+\alpha_{yy}=0\text{, and, }\forall\ k,\ \sum_{j\neq k}^{K}\alpha_{kj}+\alpha_{ky}=0\quad. \tag{9}$$

The Unrestricted Model

The unrestricted model incorporates the additional possibility that the firm fails to achieve allocative efficiency. To implement a test for allocative inefficiency, we adopt the shadow-price approach introduced by Lau and Yotopoulos (1971), used by Atkinson and Halvorsen (1980), and extended by Atkinson and Cornwell (1994). This approach ensures that the first-order conditions for profit maximization are satisfied by defining a vector of shadow input prices $w^S = \theta w$, where θ is a vector that distorts the market prices w, leading to the first-order conditions

$$w_k^S = \left(pe^u \right) \frac{\partial F}{\partial v_k} \qquad \forall\ k. \tag{10}$$

If $\theta = 1$, then the shadow prices for the inputs are the same as their market prices, and there is no allocative inefficiency. However, if $\theta \neq 1$ then the shadow-price vector differs from the market-price vector, and the firm employs an allocatively inefficient input mix. Focusing on the employment of labor (factor 1), if $\theta_1 < 1$ then the firm over-employs labor because it behaves as if labor's wage were lower than it is. Conversely, if $\theta_1 > 1$ then the firm under-employs labor. Thus, empirical estimates of θ_1 provide evidence that can be used to test the opposing predictions of the Stole-Zwiebel and de Fontenay-Gans hypotheses.

The shadow profit function is defined by

$$\pi^S(pe^u, w^S) \equiv \underset{v}{Max} \left\{ \left(pe^u \right) y - w^S \cdot v;\ y = F(v) \right\}, \tag{11}$$

and is related to actual profit π^a in the following way:

$$\begin{aligned} \pi^a &= \left(pe^u \right) F(v) - \sum_{k=1}^{K} w_k v_k \\ &= \pi^S(pe^u, w^S) - \sum_{k=1}^{K} (w_k - w_k{}^S) v_k \\ &= \pi^S(pe^u, w^S) \left[1 - \sum_{k=1}^{K} \frac{(1-\theta_k)}{\theta_k} S_k{}^S \right], \end{aligned} \tag{12}$$

where $S_k{}^S = -\dfrac{\partial \ln \pi^S(pe^u, w^S)}{\partial \ln w_k{}^S}$ is the shadow cost share for the *k*-th input. The actual cost share for the *k*-th input is $\dfrac{w_k v_k}{\pi^a}$. To estimate (12), we adopt the translog form for the shadow profit function,

$$\ln \pi^a = \ln \pi^S(pe^u, w^S) + \ln\left[1 - \sum_{k=1}^{K} \frac{(1-\theta_k)}{\theta_k} S_k{}^S\right], \tag{13a}$$

and estimate the system of K + 1 equations given by (13a), together with the actual cost-share equations

$$S_k{}^a \equiv \frac{w_k v_k}{\pi^a} = \left[1 - \sum_{k=1}^{K} \frac{(1-\theta_k)}{\theta_k} S_k{}^S\right]^{-1} \left(S_k{}^S \frac{1}{\theta_k}\right) \qquad \forall\ k, \tag{13b}$$

while imposing the parameter restrictions indicated at (9), where $\ln \pi^S(pe^u, w^S)$ and $S_k{}^S$ are given by (8a) and (8b), respectively, with w^S replacing w.

A Test of the Size of the Front-load Factor

If empirical estimates of the restricted and unrestricted models reveal no evidence of allocative inefficiency, then neither the Fontenay-Gans nor the Stole-Zwiebel model is consistent with our data. However, if the unrestricted model better explains our data and the estimate for θ_1 is greater than one, then the evidence would be consistent with the under-employment of labor predicted by the Fontenay-Gans model, leading us to reject the Stole-Zwiebel hypothesis as an explanation for the pattern of employment in our sample. On the other hand, an estimate for θ_1 of less than one would be consistent with the over-employment predicted by the Stole-Zwiebel model, leading us to reject the Fontenay-Gans hypothesis. A finding over-employment of labor, however, is not conclusive evidence in support of the Stole-Zwiebel hypothesis, since their model also predicts that all firms' front-load factors are equal to one.

Therefore, if evidence of over-employment of labor is found, our estimation procedure calls for a second step in which the front-load factors are estimated for the firms in our sample. Since the front-load factor is the ratio of Stole-Zwiebel profit to neoclassical profit, and the former is based on the sum of neoclassical profits earned over all inframarginal employment levels, estimates of the unrestricted model can be used in this second step to obtain predicted values for neoclassical profit at alternative levels of employment, allowing us to calculate each firm's front-load factor.

Construction of the Stole-Zwiebel profit level for a firm with *n* employees involves summing the predicted neoclassical profit associated with *i* employees for $i = 0, 1, \ldots, n$. Since our sample does not contain firms with every possible integer number of employees represented, we estimate the relationship between predicted neoclassical profit and the level of employment for each firm. This relation is specified as

$$\ln \pi_j = \beta_1 emp_j + \beta_2 emp_j^2 + \beta_3 TE_j + \psi Z_j + \varepsilon_j, \tag{14}$$

where π_j denotes the predicted neoclassical profit of firm *j* obtained from the estimated profit equation (13a), emp_j denotes the number of employees at firm *j*, $TE_j = e^{-u_j}$ denotes the *j*-th firm's technical inefficiency, which can be viewed as a proxy for the firm's choice of technology, Z_j is a vector of control variables representing quasi-fixed inputs, ψ is a vector of corresponding parameters, and ε_j is a normally distributed error term. Once this relationship is estimated, we can calculate neoclassical profit for a firm with any number *i* of employees. Using equation (2b), we then calculate the Stole-Zwiebel profit for a firm with *n* employees and, using equation (6b), determine each firm's front-load factor using equation (6b). Given that $\gamma(F,n)=1$ is equivalent to $\pi(n)=\tilde{\pi}(n)$, the null hypothesis for an individual firm *j* with *n* employees is

$$H_O : \delta_j \equiv \ln \gamma_j \equiv \ln \tilde{\pi}_j(n) - \ln \pi_j(n) = 0. \tag{15}$$

A distributional assumption about the estimated front-load factor is crucial to testing this hypothesis. Without such an assumption, the estimated correlation could not provide a measure of how close the two series, $\ln \pi(n)$ and $\ln \tilde{\pi}(n)$, are. Accordingly, we appeal to the assumption that the error term in equation (14) is normally distributed, which implies that $\ln \hat{\pi}_j(n)$ follows a normal distribution, as does $\ln \hat{\tilde{\pi}}_j(n)$, because the sum of normally distributed random variables is also normally distributed. Then, the vector $\hat{\delta}$ containing estimated values of δ_j for each firm follows a normal distribution with mean $\mu = E(\hat{\delta})$ and covariance $\Sigma = Cov(\hat{\delta})$. Derivations of the mean and the covariance of $\hat{\delta}$ are provided in Appendix 1.

The hypothesis in (15) states that the ratio of Stole-Zwiebel profit to neoclassical profit (or the front-load factor) of an individual firm equals one. However, there are insufficient data in our sample to carry out this test. Instead, we perform our test at the industry level so that testing the null hypothesis

$$H_O : \delta_1 = 0, \ \cdots, \ \delta_j = 0, \ \cdots, \ \delta_N = 0 \tag{16}$$

entails a joint test of the hypothesis that all firms' front-load factors are unitary for the *N* firms in our sample. Under the null hypothesis in (16), the test statistic

$$\hat{F} = \hat{\delta}^{\hat{T}} \hat{\Sigma}^{-1} \delta \tag{17}$$

follows an F distribution with $N-1$ and $N-K-1$ degrees of freedom, where $\hat{\delta}^{T} = [\delta_1, \cdots, \delta_N]$. A derivation of the distribution of the test statistic is provided in Appendix 2.

4. Data and Estimation Procedure

The Data

We implement the tests developed in the previous section using panel data on Korean savings banks. Korean savings banks were first established as local financial institutions designed to provide more convenient financial services for working-class individuals and small-and-medium-sized enterprises, and began legal operations in 1972. Prior to 1972, savings banks existed but were not legally authorized to lend money. Nevertheless, they made available various types of private loan services to borrowers who did not have access to commercial banks because of insufficient collateral, no credible source of income, or a bad credit history. In providing these loan services, savings banks developed a comparative advantage in collecting otherwise unobservable information on the relevant credit risks from its past and ongoing relationships with these sub-prime borrowers. Savings banks began to compete directly with commercial banks for deposits and loans after the financial deregulation of the late 1980s and early 1990s.[2] However, savings banks' tradition of providing loans to relatively small, risky borrowers based upon past relationships with such borrowers persists to the present.

The banking industry possesses two unique characteristics that facilitate a test of the Stole-Zwiebel hypothesis. First, banks are subject to regulatory supervision which

[2] The full range of payment settlements and foreign-currency services is performed only by commercial banks, however.

generates detailed information on their profits and asset portfolios. Second, a bank is a financial intermediary which transforms various financial and physical resources into loans and investments. As Sealey and Lindley (1977) pointed out, the failure to consider the intermediation function of banks has led researchers to misidentify outputs and inputs and to analyze incorrectly the technical aspects of production and cost in the banking industry.

Within this intermediation approach, the relevant variables are defined in the following way. First, the variable inputs are labor (v_1) and borrowed money (v_2). The corresponding prices of the variable inputs are constructed as follows: the price of labor (w_1) is the sum of salaries and employment benefits divided by the number of employees; and the price of borrowed money (w_2) is the interest paid on borrowed money divided by the volume of borrowed money. Second, we follow Berger and Mester (1997) and introduce the quasi-fixed inputs they suggest to control for special characteristics of the banking industry: off-balance sheet items (z_1) to control for the quality of loans, on the assumption that credit risks increase as the size of loans increase;[3] financial capital (z_2) to control for regulatory supervision, on the assumption that banks must meet regulatory capital requirements; and physical capital (z_3) to circumvent the difficulty of measuring its price. Third, the output variable is defined to be bank revenues (y), and output price (p) is calculated as operating revenue divided by total assets net of physical

---

[3] Other studies have used different variables to control for the quality of loans. For example, Hughes and Mester (1993) and Mester (1996) used non-performing loans, while Berg et al. (1992) used loan losses.

capital. Finally, variable profit (π^a) is defined as operating revenue net of variable costs. Accounting for the quasi-fixed inputs, the short-run variable-profit function is

$$\pi(pe^u, w, z) \equiv \max_v \left\{ \left(pe^u \right) y - w \cdot v;\; y = F(v, z) \right\}. \tag{18}$$

The empirical profit and cost-share equations (8a)-(8b) and (13a)-(13b) are also modified to include the quasi-fixed inputs.

The data are taken from information collected by the Korean Financial Supervisory Service (FSS). All Korean savings banks are required to submit annual reports to the FSS. We use annual data reported each June for the years 2002 through 2008. Since we are estimating a translog profit function, our sample is limited to the 42 banks reporting positive profit every year.[4] All data are deflated by the GDP deflator, as is customary in this literature [e.g. Berger and Mester (1997) and Wheelock and Wilson (1999)]. Table 2 provides summary statistics for the variables describing the savings banks in the sample.

Estimation Procedure

When panel data are available, the fixed-effects estimator is commonly used to identify unobserved firm-specific effects with technical inefficiency [Schmidt and Sickles (1984)].[5] With fixed-effects estimation, allocative inefficiency can be parameterized as Atkinson and Cornwell (1994) suggested in their estimation of a cost system. However,

---

[4] If a single profit equation were being estimated, then the rescaling method (making negative profits positive by adding a constant to the negative profits) or the indicator method (making negative profits equal to 1 by adding an indicator variable to the left-hand side) could be utilized [Bos and Koetter (2011]. However, neither method is appropriate when estimating a system of profit and cost-share equations because the rescaling method distorts revenue and cost shares and the indicator method does not properly rescale the share equations.

[5] If only cross-section data are available, the stochastic frontier estimator is most commonly used [Aigner, Lovell and Schmidt (1977)].

it is extremely difficult computationally to separate firm-specific, unobserved heterogeneity from the other explanatory variables in the system of profit and cost-share equations in (13), because technical inefficiency u is embedded in the profit function. This computational difficulty explains why such systems of equations embodying technical and allocative inefficiencies have not been successfully estimated with panel data.

We avoid this computational difficulty by imposing the assumption that the production process is homogeneous of degree $r < 1$ in the variable inputs. We exploit this assumption in our estimation procedure by taking advantage of the following definition and propositions.[6]

Definition: The normalized variable profit function, $\hat{\pi}(w,z)$, is defined as

$$\hat{\pi}(w,z) \equiv \underset{v}{Max}\{F(v,z) - \sum_{k=1}^{K} w_k v_k\}.$$

Proposition 1: $\hat{\pi}(cw,z) = c^{\frac{r}{r-1}}\hat{\pi}(w,z)$ if and only if $F(v,z)$ is homogeneous of degree r in v, where c > 0 is a scalar.

Proposition 2: $\pi(pe^{u},w,z) = \left(pe^{u}\right)^{\frac{1}{1-r}}\hat{\pi}(w,z)$ if and only if $F(v,z)$ is homogeneous of degree r in v.

[6] Propositions 1 and 2 were established by Lau (1978); proofs of Propositions 1-3 can be found in Lau (1978) and Kumbhakar (2001).

Proposition 3: Linear homogeneity of $\pi(pe^{u}, w, z)$ in (pe^{u}, w) is equivalent to $\hat{\pi}(w,z)$ being homogeneous of degree $\frac{r}{r-1}$ in w if and only if $F(v,z)$ is homogeneous of degree r in v.

When the production technology is homogenous of degree r in the variable inputs, Proposition 1 identifies the homogeneous structure of the normalized profit function, Proposition 2 states that the technical-efficiency-adjusted output price can be separated from the profit function by a factor multiplying the normalized profit function, and Proposition 3 asserts that the linear homogeneity property of the profit function is equivalent to homogeneity of degree $r/(r-1)$ for the normalized profit function. Using these results, equation (13a), with the quasi-fixed inputs included, can be re-specified as

$$\ln \pi^{a} = \alpha_0 \ln p + \ln \hat{\pi}^{S}(w^{S}, z) + \ln\left[1 - \sum_{k=1}^{K} \frac{(1-\theta_k)}{\theta_k} S_k{}^{S}\right] + u + \varepsilon, \tag{19a}$$

where $\alpha_0 = 1/(1-r)$, $S_k^S = -\alpha_k - \sum_{j=1}^{3} \alpha_{kj} \ln w_j^S$, and

$$\begin{aligned}\ln \hat{\pi}^{S}(w^{S}, z) = &\sum_{k=1}^{2} \alpha_k \ln w_k^S + \sum_{m=1}^{3} \beta_m \ln z_m \\ &+ \frac{1}{2}\left[\sum_{k=1}^{2}\sum_{j=1}^{2} \alpha_{kj} \ln w_k^S \ln w_j^S + \sum_{m=1}^{3}\sum_{l=1}^{3} \beta_{ml} \ln z_m \ln z_l\right] \\ &+ \sum_{k=1}^{2}\sum_{m=1}^{3} \lambda_{km} \ln w_k^S \ln z_m\end{aligned} \tag{19b}$$

with the symmetry restrictions, $\alpha_{kj} = \alpha_{jk}$ and $\beta_{ml} = \beta_{lm}$ imposed. Lau (1978, p. 131) showed that the normalized profit function is monotonically nonincreasing and convex in w. The cost-share equations are given by (13b) with $K = 2$ and $S_k^S = -\alpha_k - \sum_{j=1}^{2} \alpha_{kj} \ln w_j^S$.

Finally, Proposition 3 indicates that the normalized profit function is homogeneous of degree $(1-\alpha_0)$, implying the parameter restrictions

$$\alpha_0 + \sum_{k=1}^{2}\alpha_k = 1, \sum_{j=1}^{2}\alpha_{kj} = 0 \text{ for } k = 1, 2,$$
$$\sum_{k=1}^{2}\lambda_{km} = 0 \text{ for } m = 1, 2, 3, \text{ and } \sum_{k=1}^{2}\sum_{j=1}^{2}\alpha_{kj} = 0. \quad (20)$$

When quasi-fixed inputs are included, the system of equations in (8a) and (8b) is given by (19a), (19b), and (13b), with θ_k set equal to one for all *k* to yield a profit system with only technical inefficiency.

Technical inefficiency u is specified to have the time-varying form

$$\mu_{it} = \mu_i + \mu^1 t + \mu^2 t^2 \quad (21)$$

where μ_i is a time-invariant, bank-specific parameter, and μ^1 and μ^2 are parameters that are common to all banks. It is possible to specify more flexible forms of technical inefficiency or to disentangle inefficiency and other forms of heterogeneity; however, we use this more parsimonious specification to conserve degrees of freedom.[7] Since u is allowed to be time-varying, we calculate technical inefficiency relative to the best-performing bank over the entire sample period.

We estimate the equation system by iterative feasible generalized least squares, and calculate heteroskedasticity-consistent (robust) standard errors. The allocative inefficiency terms are estimated as an industry mean because the short panel dataset is subject to degrees-of-freedom limitations. As suggested by Cornwell et al. (1990), the

[7] For example, as in Atkinson and Primont (2002), technical inefficiency could be specified as $u_{it} = u_i + u_i^1 t + u_i^2 t^2$ where u_i, u_i^1, and u_i^2 are firm-specific parameters. However, this specification requires the estimation of $2(N-2)$ additional parameters, where N is the number of firms. Similarly, in Greene's (2005) "true" fixed-effects model the firm-specific error is specified as $v_{it} = a_i + u_{it}$, where a_i is unobserved heterogeneity and u_{it} is time-varying technical inefficiency, requiring the estimation of $N-2$ additional parameters.

estimates of $u+\varepsilon$ are then regressed on the right-hand side variables in (21) to obtain time-varying, firm-specific measures of technical inefficiency under the assumption that the coefficient estimates of the profit system are consistent.[8]

5. Empirical Results

The results of estimating the system of equations in (19a), (19b), and (13b) for the restricted model (where $\theta_k = 1 \; \forall k$) and for the unrestricted model (where θ_k is estimated) are reported in Table 3. The estimated coefficients on the explanatory variables are difficult to interpret, owing to the nonlinearity of the translog functional form. However, both monotonicity and convexity of the profit function are satisfied at the sample mean (and median) values of the variables.[9]

It is expected that the presence of allocative inefficiencies will affect the estimates of all the coefficients, but especially the estimates of the coefficients on the input prices because those allocative inefficiencies are parameterized. It is also expected that the allocative inefficiency associated with borrowed money will be smaller than that for labor because savings banks compete with commercial banks for deposits. These *a priori* expectations are realized: the coefficient estimates are quite different between the two models, and the estimate of the allocative inefficiency for borrowed money is smaller than for labor. However, the magnitude of this allocative inefficiency is small, which

[8] Wang and Schmidt (2002) and Alvarez et al. (2006) show that a two-step approach, in which estimates of technical inefficiency (u_i) obtained from a first-step stochastic-frontier estimation is regressed on "exogenous" variables such as a firm size leads to biased estimates. Their argument is that such exogenous influences must be accounted for in the first-step estimation if technical inefficiency is affected by the exogenous variables. However, this problem does not arise in our model because we use the fixed-effects estimator for which a distributional assumption on technical inefficiency is not required. For details, see Greene (2008, pp. 155-156).

[9] For the restricted model, 96% of the observations satisfy monotonicity and 92.2% of the observations satisfy convexity. For the unrestricted model, all of the observations satisfy both monotonicity and convexity.

accords with the findings of Kumbhakar and Tsionas (2005, p 378). Note that the estimated coefficient on output price is greater than one in both models. This result is consistent with our assumption that the technology is homogeneous of degree $r < 1$; that is, production is subject to decreasing returns to scale in the variable inputs.

Using these estimates, we test the null hypothesis

$$H_0 : \theta_1 = \theta_2 = 1 \tag{22}$$

(implying that allocative inefficiency is absent) against the two-sided alternative that at least one of the $\theta_k \neq 1$ (k = 1, 2). The log-likelihood-ratio (LR) statistic is

$$LR = -2(LogL_R - LogL_U) \tag{23}$$

where $LogL_R$ is the maximum value of the likelihood function for the restricted model, and $LogL_U$ is the maximum value of the likelihood function for the unrestricted model. LR follows a χ^2 distribution with two degrees of freedom (equal to the number of restrictions imposed). The χ^2 critical value at the 0.01 significance level is 9.210 and the LR statistic is 271.4. Thus, the null hypothesis is strongly rejected, and we conclude that Korean savings banks exhibit allocative inefficiency. Specifically, the estimate of θ_1 is 0.5137 < 1 ($p < 0.01$), indicating that the shadow wage rate is 51 percent of the actual wage rate, which is consistent with the Stole-Zwiebel prediction of over-employment rather than the de Fontenay-Gans prediction of under-employment. [10] As shown in Figure 2, Korean savings banks are also technically inefficient. Despite moderate

[10] Using the estimated value of θ_2 and its respective standard error from the unrestricted model, reported in Table 3, we also reject the null hypothesis $\theta_2 = 0$ in favor of $\theta_2 < 1$, with a p-value of less than 0.01.

improvement over time, the magnitude of this technical inefficiency is large.[11] These allocative and technical inefficiencies may help explain why savings banks in Korea are less profitable than commercial banks.

Given our finding of allocative inefficiency consistent with over-employment, we then estimate the relationship between predicted neoclassical profit and the level of employment specified in (14). We average over the sample period each firm's predicted neoclassical profit, number of employees, level of technical (in)efficiency, and levels of the quasi-fixed inputs, and then perform OLS (i.e., "between") estimation using the firm means of these variables. The estimated regression is

$$\begin{aligned}\ln \hat{\pi}_{j.} = &\underset{(0.196)}{6.438} + \underset{(0.0061)}{0.011}\, emp_{j.} - \underset{(0.000027)}{0.000027}\, emp_{j.}^2 + \underset{(0.3965)}{1.4052}\, TE_{j.} \\ &+ \underset{(0.0089)}{0.0105}\, z_{1,j.} + \underset{(0.0028)}{0.0208}\, z_{2,j.} - \underset{(0.0068)}{0.0032}\, z_{3,j.} \qquad (24)\end{aligned}$$

$$N = 45 \quad \bar{R}^2 = 0.845$$

where robust standard errors are in parentheses. These results identify a (weakly) quadratic relation between predicted neoclassical profit and the level of employment, controlling for the choice of technology.[12] This concave relation is consistent with the predictions of the Stole-Zwiebel model. Using this empirical relation between neoclassical profit and the level of employment, each firm's front-load factor is

[11] Berger and Humphrey (1997) concluded from their survey of the relevant literature that the mean technical inefficiency for U.S. banks is 16%.

[12] In addition, both fixed-effects and random-effects estimates of (24), which are not reported, capture this quadratic relation.

constructed using the current level of employment and the industry-wide average level of technical efficiency. The results are shown in Figure 3.[13]

We test the null hypothesis in (16) that all of these front-load factors equal one. The test statistic, which under the null hypothesis follows an F distribution with $N-1=41$ and $N-K-1=35$ degrees of freedom, is 0.117. The critical value for the F statistic at the 0.05 level of significance is 1.73. Thus, the null hypothesis that all of the front-load factors are unity cannot be rejected, supporting the Stole-Zwiebel prediction that firms hire workers until the bargained wage equals the competitive-market wage.

6. Conclusions

Stole and Zwiebel argue that firms respond to the employee hold-up power implicit in relationship-specific human capital by over-employing labor, leading to allocative inefficiency. Technical inefficiency is a potential by-product of this over-employment since firms are willing to sacrifice productive efficiency if that will enhance their profitability, given the wage-bargaining process. These input distortions can be characterized by a single statistic, the front-load factor, which is based on the neoclassical profit that would be earned at each level of employment up to the actual level. Firms are predicted to expand employment until the bargaining-determined wage is driven down to the market-determined wage, where the front-load factor is equal to one.

We develop and implement a two-step procedure for testing these two key predictions of the Stole-Zwiebel model of intra-firm wage bargaining. The first step in our approach tests for the presence of allocative inefficiency. Using panel data on Korean savings banks, we estimate a translog system of profit and cost-share equations. Our first

[13] By construction, each firm's front-load factor is less than or equal to one. We are interested in testing whether these factors are jointly equal to one against the alternative that at least one of them is less than one.

set of empirical results provides evidence of allocative inefficiency in the form of over-employment. We then determine each bank's front-load factor by estimating the relation between the predicted neoclassical profit and observed employment. Based on these results, we cannot reject the null hypothesis that every firm's front-load factor is equal to one. From this evidence, we conclude that the employment decisions of Korean savings banks are consistent with two key predictions of the Stole-Zwiebel model of intra-firm wage bargaining. Our findings help explain why these savings banks employ more workers per branch than their commercial-bank counterparts, despite the smaller contribution to profits of the average savings-bank employee.

Empirical studies of the banking sector typically find evidence of substantial allocative and technical inefficiencies. Some of the policy recommendations stemming from the presence of these distortions focus on changing regulations that affect industry competitiveness (such as barriers to entry and guidelines for mergers and acquisitions), the market for corporate control, and the scope of permissible lines of business. By contrast, the Stole-Zwiebel model provides an endogenous explanation for technical inefficiency that arises as a response to the over-employment associated with the prospect of hold up by employees with private knowledge that is valuable to the firm. In Korean savings banks, loan officers with special knowledge of borrowers' risk characteristics present such a hold-up threat. Our interpretation of the empirical findings suggest that the allocative and technical inefficiency exhibited by Korean savings banks might be reduced by requiring more transparent accounting practices by the small businesses that comprise the majority of their borrowers, thereby reducing the potential hold-up power of loan officers in these banks.

Appendix 1

In this appendix we derive the mean and covariance for the estimate of the vector δ. Let the employment equation (14) be rewritten, for simplicity, as

$$Y = XB + \varepsilon \tag{A1-1}$$

where $Y = \log \pi$, $X = [emp, emp^2, TE, z_1, z_2, z_3]$ and $B = [\beta_1, \beta_2, \beta_3, \psi_1, \psi_2, \psi_3]$. Then, for a firm's time-mean frontload factor, (A1-1) can be expressed as

$$y_{j.} = x_{j.}^{T} \beta + \alpha + \varepsilon_{j.} \tag{A1-2}$$

where $y_{j.} = \sum_{t=1}^{T} y_{jt} / T$, $x_{j.} = \sum_{t=1}^{T} x_{jt} / T$, and $\varepsilon_{j.} = \sum_{t=1}^{T} \varepsilon_{jt} / T$ so that between estimation is applied. With (A1-2), we can construct the time-mean neoclassical profit of firm *j* at the current level of employment n_j and technical efficiency *TE* as

$$y_{j.}(n_{j.}; TE, Z) = \left(x_{j.}\right)^{T} \hat{\beta} + \hat{\alpha} + \varepsilon_{j.}\,. \tag{A1-3}$$

Under the assumption $\varepsilon_{jt} \sim N(0, \sigma_{\varepsilon}^{2} I_T)$ for all *j*, where I_T is a (*T x T*) identity matrix, the mean and variance of the neoclassical profit of firm *j* is

$$E\left(y_{j.}(n_{j.}; TE_{j.}, Z_{j.})\right) = \left(x_{j.}\right)^{T} \beta + \alpha$$
$$Var\left(y_{j.}(n_{j.}; TE_{j.}, Z_{j.})\right) = Var\left(\frac{\sum_{t=1}^{T} y_{jt}(n_{jt}; TE_{jt}, Z_{jt})}{T}\right) = \frac{\sigma_{\varepsilon}^{2}}{T}. \tag{A1-4}$$

Following (2b), the time-mean Stole-Zwiebel profit of firm *j* at the current level of employment n_j and technical efficiency *TE* is calculated as

$$\tilde{y}_{j.}(n_{j.};TE_{j.},Z_{j.}) = \frac{\sum_{i=1}^{n_{j.}} y_{j.}(i;TE_{j.},Z_{j.})}{n_{j.}+1} = \frac{1}{(n_{j.}+1)T}\sum_{i=1}^{n_{j.}}\sum_{t=1}^{T} y_{jt}(n_{jt};TE_{jt},Z_{jt}) \tag{A1-5}$$

and the mean and the variance of Stole-Zwiebel profit for firm *j* are thus given by

$$\begin{aligned} E\left(\tilde{y}_{j.}(n_{j.};TE_{j.},Z_{j.})\right) &= \frac{1}{(n_{j.}+1)T}\sum_{i=1}^{n_{j.}}\sum_{t=1}^{T}\left(x_{jt}\right)^T \beta + \alpha \\ Var\left(\tilde{y}_{j.}(n_{j.};TE_{j.},Z_{j.})\right) &= Var\left(\frac{1}{(n_{j.}+1)T}\sum_{i=1}^{n_{j.}}\sum_{t=1}^{T} y_{jt}(n_{jt};TE_{jt},Z_{jt})\right) \\ &= \frac{n_{j.}}{(n_{j.}+1)^2}\frac{\sigma_\varepsilon^2}{T}. \end{aligned} \tag{A1-6}$$

Therefore, the logarithm of firm *j*'s time-mean front-load factor $(\delta_{j.})$ at the current level of employment n_j and technical inefficiency *TE* has the following mean and variance:

$$\begin{aligned} E(\hat{\delta}_{j.}) &= \left(x_{j.}\right)^T \beta - \frac{1}{(n_{j.}+1)T}\sum_{i=1}^{n_{j.}}\sum_{t=1}^{T}\left(x_{jt}\right)^T \beta \\ Var(\hat{\delta}_{j.}) &= \frac{\sigma_\varepsilon^2}{T}\left(\frac{n_{j.}{}^2 + n_{j.} + 1}{n_{j.}{}^2 + 2n_{j.} + 1}\right). \end{aligned} \tag{A1-7}$$

Assuming that one firm's behavior is independent of the other firms' behavior, the mean and covariance of $\hat{\delta}$ are, respectively,

$$E(\hat{\delta}) = E\begin{bmatrix} \hat{\delta}_1 \\ \vdots \\ \vdots \\ \hat{\delta}_N \end{bmatrix}$$

and

$$Cov(\hat{\delta}) = \begin{bmatrix} Var(\hat{\delta}_1) & \cdots & Cov(\hat{\delta}_1, \hat{\delta}_N) \\ \vdots & \ddots & \vdots \\ Cov(\hat{\delta}_N, \hat{\delta}_1) & \cdots & Var(\hat{\delta}_N) \end{bmatrix}$$

$$= \sigma_\varepsilon^{\,2} \begin{bmatrix} \frac{1}{T}\frac{n_{1.}^{\,2}+n_{1.}+1}{n_{1.}^{\,2}+2n_{1.}+1} & 0 & \cdots & 0 \\ 0 & \frac{1}{T}\frac{n_{j.}^{\,2}+n_{j.}+1}{n_{j.}^{\,2}+2n_{j.}+1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & \frac{1}{T}\frac{n_{N.}^{\,2}+n_{N.}+1}{n_{N.}^{\,2}+2n_{N.}+1} \end{bmatrix} \qquad \text{(A1-8)}$$

$$= \sigma_\varepsilon^{\,2}\Omega = \Sigma.$$

Appendix 2

In this appendix we derive the distribution of the test statistic $\hat{F}$ in (17). We assume that the distribution of $\hat{\delta}$ is normal. Then, under the assumption that σ_ε^2 is known, the test statistic $\hat{F}$ follows a non-central χ^2 with rank $N-1$ where N is the number of firms in our sample and non-centrality parameter $\lambda = \frac{1}{2}\delta^T \Sigma^{-1}\delta$, so that

$$\hat{F} = \hat{\delta}^T \Sigma^{-1}\delta \sim \text{non-central } \chi^2_{r,\lambda}. \qquad \text{(A2-1)}$$

However, since σ_ε^2 is unknown, we use the estimated value from the between panel estimation, $\hat{\sigma}_\varepsilon^{\,2} = \frac{\left(\hat{\varepsilon}_{j.}\right)^T \varepsilon_{j.}}{N-K-1}$, where K = 6 and $\hat{\varepsilon}_{j.}$ is a vector of residuals obtained from (A1-2). Hence,

$$\frac{\hat{\sigma}_\varepsilon^{\,2}(N-K-1)}{\sigma_\varepsilon^{\,2}} \sim \chi^2_{N-K-1}. \qquad \text{(A2-2)}$$

Using (A2-2), the test statistic $\hat{F}$ can be expressed as the ratio of two χ^2 distributions:

$$\frac{\hat{F}}{N-1} = \left(\frac{\hat{\delta}^{\mathrm{T}}\Omega^{-1}\delta}{\hat{\sigma}_{\varepsilon}^{2}} \right) / (N-1) = \frac{\left(\frac{\hat{\delta}^{\mathrm{T}}\Omega^{-1}\delta}{\sigma_{\varepsilon}^{2}} \right) / (N-1)}{\left(\frac{\hat{\sigma}_{\varepsilon}^{2}(N-K-1)}{\sigma_{\varepsilon}^{2}} \right) / (N-K-1)}. \tag{A2-3}$$

Finally, since the ratio of two χ^2 distributions follows an F distribution, (A2-3) follows a central F distribution with $N-1$ and $N-K-1$ degrees of freedom under the null hypothesis H_0 stated at (16). That is,

$$\frac{\hat{F}}{N-1} \sim \text{central } F(N-1, N-K-1). \tag{A2-4}$$

References

Aigner DJ, Lovell CAK, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 6(1): 21-37.

Alvarez A, Amsler C, Orea L, Schmidt P (2006) Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics. Journal of Productivity Analysis 25(3): 201-212.

Atkinson SE, Halvorsen R (1980) A test of relative and absolute price efficiency in regulated utilities. Review of Economics & Statistics 62(1): 81-88.

Atkinson SE, Cornwell C (1994) Parametric estimation of technical and allocative inefficiency with panel data. International Economic Review 35(1): 231-243.

Atkinson SE, Primont D (2002) Stochastic estimation of firm technology, inefficiency, and productivity growth using shadow cost and distance functions. Journal of Econometrics 108(2): 203-225.

Berg SA, Forsund FR, Jansen ES (1992) Malmquist indices of productivity growth during the deregulation of Norwegian banking. Scandinavian Journal of Economics 94: S211-S228.

Berger AN, Humphrey D (1997) Efficiency of financial institutions: international survey and directions for future research. European Journal of Operations Research 98(2): 175-212.

Berger AN, Mester LJ (1997) Inside the black box: what explains differences in the efficiencies of financial institutions? Journal of Banking and Finance 21(7): 895-947.

Bertola, P, Garibaldi, G (2001) Wages and the size of firms in dynamic matching models. Review of Economic Dynamics 4(2): 335-368.

Bos JWB, Koetter M (2011) Handling losses in translog profit models. Applied Economics 43(3): 307-312.

Cahuc, P, Wasmer E. (2001). Does intrafirm bargaining matter in the large firm's matching model? Macroeconomic Dynamics 5(5): 742-747.

Chakrabarti, S, Tangsangasaksri, S (2011) Labor unions and Nash bargaining using coalition formation games. Mathematical Social Sciences 62(1): 55-64.

Cornwell, C, Schmidt P, Sickles RC (1990) Production frontiers with cross-sectional and time-series variation in efficiency levels. Journal of Econometrics 46(1-2): 185-200.

De Fontenay CC, Gans JS (2003) Organizational design and technology choice under intrafirm bargaining: comment. American Economic Review 93(1): 448-455.

Greene WH (2005) Fixed and random effects in stochastic frontier models. Journal of Productivity Analysis 23(1): 7-23.

Greene WH (2008) The econometric approach to efficiency analysis. In: Fried HO, Lovell CAK, Schmidt SS (eds.) The Measurement of Productive Efficiency and Productivity Growth. Oxford University Press, Oxford, pp. 92-250.

Hughes JP, Mester LJ (1993) A quality and risk-adjusted cost function for banks: evidence on the “too-big-to-fail” doctrine. Journal of Productivity Analysis 4(3): 293-315.

Kumbhakar, SC (2001) Estimation of profit functions when profit is not maximum. American Journal of Agricultural Economics 83(1): 1-19.

Kumbhakar SC, Tsionas EG (2005) Measuring technical and allocative inefficiency in the translog cost system: a Bayesian approach. Journal of Econometrics 126(2): 355-384.

Lau LJ (1978) Applications of profit functions. In: Fuss M, McFadden D (eds) Production Economics: A Dual Approach to Theory and Applications, vol 1, North-Holland, Amsterdam, pp. 133-216.

Lau LJ, Yotopoulos PA (1971) A test for relative efficiency and application to Indian agriculture. American Economic Review 61(1): 94-109.

Malcomson, JM (1997) Contracts, hold-up, and labor markets. Journal of Economic Literature 35(4): 1916-1957.

Mester, LJ (1996) A study of bank efficiency taking into account risk-preference. Journal of Banking and Finance 20(6): 1025-1045.

Rebitzer, JB, Taylor LJ (2007) When knowledge is an asset: explaining the organizational structure of large law firms. Journal of Labor Economics 25(2): 201-229.

Schmidt P, Sickles RC (1984) Production frontiers and panel data. Journal of Business and Economics and Statistics 2(4): 367-374.

Sealey CW, Lindley JT (1977) Inputs, outputs, and a theory of production and cost at depository financial institutions. Journal of Finance 32(4): 1251-1266.

Stole LA, Zwiebel J (1996a) Organizational design and technology choice under intrafirm bargaining. American Economic Review 86(1): 195-222.

Stole LA, Zwiebel J (1996b) Intra-firm bargaining under non-binding contracts. Review of Economic Studies 63(216): 375-410.

Stole LA, Zwiebel J (2003) Organizational design and technology choice under intrafirm bargaining: reply. American Economic Review 93(1): 456-457.

Wang H, Schmidt P (2002) One-step and two-step estimation of exogenous variables on technical efficiency levels. Journal of Productivity Analysis 18(2): 129-144.

Wheelock DC, Wilson PW (1999) Technical progress, inefficiency, and productivity change in U.S. banking, 1984-1993. Journal of Money, Credit and Banking 31(2): 212-234.

Table 1: *Descriptive statistics for savings and commercial banks: 2002-2008*[1]

	2002	2003	2004	2005	2006	2007	2008
Workers per branch	**26.7**	**27.6**	**25.7**	**25.3**	**24.7**	**24.9**	**24.1**
	20.8	20.1	19.6	19.0	18.9	18.9	18.8
Profit per worker (₩ million)	**20.9**	**20.4**	**3.1**	**-53.4**	**84.2**	**87.3**	**43.7**
	35.9	38.3	13.2	70.0	105.0	98.2	107.0
Wages per worker (₩ million)	**28.7**	**33.9**	**37.8**	**40.1**	**45.7**	**49.0**	**48.5**
	39.7	46.6	53.7	60.9	67.8	69.1	72.9
Branches per bank	**2.0**	**2.0**	**2.1**	**2.2**	**2.4**	**2.7**	**3.0**
	439.9	458.6	492.5	499.3	519.8	542.7	566.1
Workers per bank	**53.6**	**55.4**	**54.5**	**56.5**	**59.3**	**67.0**	**73.1**
	6,390.8	6,318.6	6,926.8	6,766.5	6,927.8	7,305.5	7,664.8
ROA (%)	**0.6**	**0.5**	**0.1**	**-0.9**	**1.4**	**0.9**	**0.6**
	0.7	0.6	0.2	0.9	1.3	1.1	1.1

[1] The descriptive statistics for savings banks are in **bold** font.

Table 2: *Means and standard deviations of variables: 2002-2008*[1,2]

	2002	2003	2004	2005	2006	2007	2008
p	0.117	0.121	0.106	0.103	0.105	0.109	0.102
	(0.036)	(0.031)	(0.022)	(0.018)	(0.020)	(0.020)	(0.013)
w_1	0.028	0.032	0.037	0.042	0.046	0.051	0.051
	(0.009)	(0.009)	(0.020)	(0.017)	(0.020)	(0.023)	(0.025)
w_2	0.065	0.056	0.051	0.049	0.044	0.047	0.052
	(0.009)	(0.007)	(0.006)	(0.004)	(0.005)	(0.005)	(0.005)
z_1	5,779	9,383	5,644	5,619	6,196	10,376	5,959
	(9,006)	(9,789)	(8,795)	(8,423)	(9,259)	(18,239)	(9,724)
z_2	12,749	14,708	17,754	22,870	29,991	38,408	44,731
	(10,734)	(11,971)	(15,644)	(23,734)	(31,981)	(42,609)	(51,244)
z_3	7,739	8,288	8,577	8,390	8,890	10,796	9,791
	(8,590)	(11,260)	(11,138)	(10,939)	(11,188)	(12,741)	(13,138)
π^a	3,060	3,399	3,505	5,855	8,256	10,843	9,660
	(4,378)	(5,577)	(4,699)	(9,870)	(8,732)	(13,727)	(13,060)

[1] Numbers in parentheses are standard deviations.

[2] Prices are measured in percentages and other variables are measured in millions of won, where p, w_1, and w_2 denote the price of output, the price of labor and the price of borrowed money, respectively. z_1, z_2, and z_3, respectively, denote the values of off-balance-sheet items, financial capital, and physical capital, while π^a denotes variable profit.

Table 3: *Empirical results*[1]

Variables	parameters	Restricted Model		Unrestricted Model	
$\ln p$	α_0	1.4671	(0.4441)	1.1879	(0.3804)
$\ln w_1$	α_1	0.8269	(0.7243)	-0.1358	(0.0385)
$\ln w_2$	α_2	-1.2941	(0.7389)	-0.0521	(0.1800)
$\ln z_2$	β_1	0.0256	(0.0891)	-0.0740	(0.0791)
$\ln z_2$	β_2	0.6490	(0.7792)	-0.6792	(0.6612)
$\ln z_3$	β_3	1.1222	(0.6386)	1.4408	(0.5313)
$\ln w_1 \ln w_1$	α_{11}	0.0797	(0.1020)	-0.0012	(0.0047)
$\ln w_1 \ln w_2$	α_{12}	0.9370	(0.1886)	-0.0025	(0.0092)
$\ln w_2 \ln w_2$	α_{22}	0.7039	(0.5025)	1.1754	(0.4291)
$\ln z_1 \ln z_1$	β_{11}	0.0003	(0.0107)	-0.0041	(0.0103)
$\ln z_1 \ln z_2$	β_{12}	0.0048	(0.0114)	0.0162	(0.0101)
$\ln z_1 \ln z_3$	β_{13}	-0.0069	(0.0070)	-0.0067	(0.0065)
$\ln z_2 \ln z_2$	β_{22}	0.0939	(0.1105)	0.2214	(0.0905)
$\ln z_2 \ln z_3$	β_{23}	-0.0546	(0.0760)	-0.0718	(0.0573)
$\ln z_3 \ln z_3$	β_{33}	-0.0773	(0.0504)	-0.1022	(0.0444)
$\ln w_1 \ln z_1$	λ_{11}	0.0012	(0.0080)	0.0002	(0.0005)
$\ln w_1 \ln z_2$	λ_{12}	0.2350	(0.0572)	0.0121	(0.0039)
$\ln w_1 \ln z_3$	λ_{13}	-0.0473	(0.0396)	-0.0036	(0.0018)
$\ln w_2 \ln z_1$	λ_{21}	-0.0282	(0.0181)	-0.0050	(0.0130)
$\ln w_2 \ln z_2$	λ_{22}	0.0880	(0.1130)	-0.0621	(0.0399)
$\ln w_2 \ln z_3$	λ_{23}	-0.0176	(0.0976)	0.0510	(0.0375)
	θ_1	1	—	0.5137	(0.0730)
	θ_2	1	—	0.8283	(0.0492)

[1]Standard errors are in parentheses.

Figure 1: *Profit maximizing choice of labor*

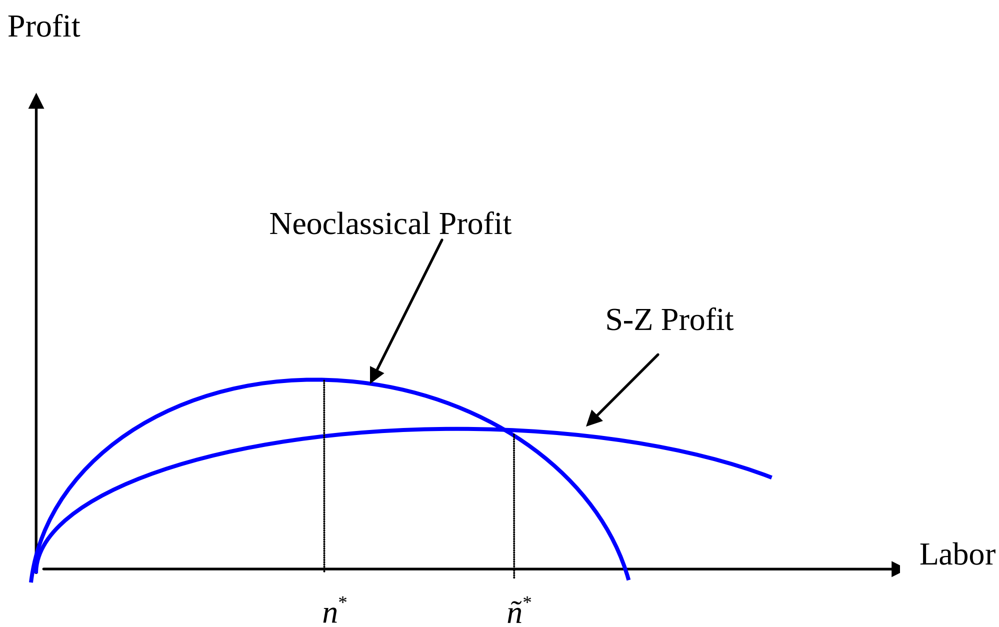

Figure 2: *Mean values and standard deviations of technical efficiency from Model A (in blue) and Model B (in red)*

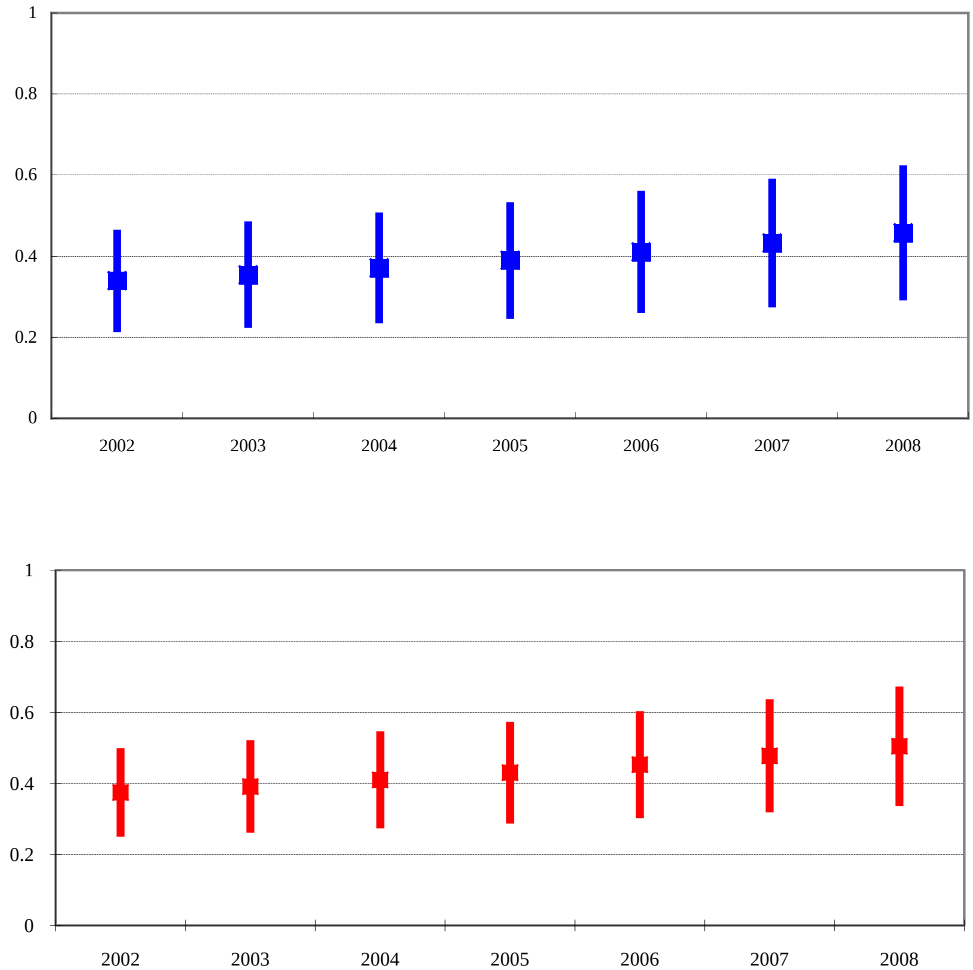

Figure 3: *Each firm's constructed front-load factor*