INTRA-FIRM WAGE BARGAINING AND ALLOCATIVE INEFFICIENCY: A TEST OF THE STOLE-ZWIEBEL HYPOTHESIS*

Yongseung Han
Gainesville State College

Arthur Snow
University of Georgia

Ronald S. Warren, Jr.*
University of Georgia

Abstract

This paper presents a test of the Stole-Zwiebel hypothesis which states that firms facing the prospect of employee hold-up power in *ex post* wage bargaining respond by over-employing labor and adopting inefficient technologies *ex ante*. We develop and implement a two-step procedure to test these predictions. The first step determines whether or not there is allocative inefficiency in the use of variable inputs. The second step determines whether the extent of the allocative inefficiency found in step one is sufficient to drive the negotiated wage down to the competitive level, as predicted by the Stole-Zwiebel model. The test is implemented using panel data on Korean savings banks. The results indicate the presence of allocative inefficiency and a degree of over-employment that is consistent with the predictions of the Stole-Zwiebel model.

JEL Classifications: D20, L20

*We thank Harold Mulherin for helpful comments on an earlier draft.

INTRA-FIRM WAGE BARGAINING AND ALLOCATIVE INEFFICIENCY: A TEST OF THE STOLE-ZWIEBEL HYPOTHESIS

I. Introduction

In the neoclassical theory of the firm, it is assumed that prices are exogenous and all factors of production are homogeneous. As a consequence, there is perfect mobility of labor and capital across firms and industries. This theory also assumes that technology is fixed, and contracts with workers, suppliers and customers are complete and costless to enforce. It has been long recognized, however, that when there are relationship-specific assets, contractual contingencies that cannot be fully specified, and potential traders are unable to commit credibly to forego contract renegotiation, *ex-post* quasi rents can be expropriated by one of the parties to a transaction. This form of opportunistic behavior, known as the "hold-up problem," may lead to *ex ante* investments in relationship-specific capital that differ from the efficient (i.e., joint-surplus-maximizing) level.

A variety of organizational arrangements (e.g., vertically integrated firms), contract provisions (e.g., non-compete clauses), and legal restrictions (e.g., resale-price maintenance laws) have been devised to reduce the scope for opportunistic behavior in the face of asset specificity, incomplete contracts, and renegotiation risks. Transactions-cost theory, initiated by Coase (1937) and extended by Williamson (1975, 1985), Klein et al. (1978), and many others, seeks to connect observed variations in such institutional arrangements with differences in the cost of mitigating inefficiencies arising from post-contractural opportunistic behavior. The property-rights or incomplete-contracts theory, associated most prominently with Grossman and Hart (1986) and Hart and Moore (1988, 1990), focuses on how, in environments where enforceable contracts cannot be written to cover all contingencies, the allocation of asset ownership affects the residual claims to the surplus created by efficient transactions, and how this, in turn, requires modifications to *ex ante* agreements between parties. Both approaches highlight the importance of resource-allocation distortions arising from attempts by traders to reduce in advance their exposure to subsequent hold up [Whinston (2003)].

Many empirical tests of the implications of transactions-cost theory have been carried out, with the evidence generally supporting the predicted relationship between the degree of asset specificity and the type of procurement contract or organizational form observed [e.g., Joskow (1987) and Nickerson and Silverman (2003), respectively]. Direct tests of the existence or extent of the hold-up problem in this literature are much less common, however, presumably because parties to transactions where hold up might occur are aware of the situation and have already adopted contracts, institutions, or policies to eliminate or reduce the expropriation of quasi-rents [Vukina and Leegomonchai (2006)]. Similarly, direct tests of the predictions of implicit-contract theory are rare because the requisite transaction-level data on the *ex ante* bargaining environment, the assignment of property rights over relationship-specific assets, and the *ex post* distribution of the surplus arising from observed transactions are generally unavailable [Whinston (2001)].

In a novel extension of these ideas, Stole and Zweibel (1996a, 1996b) emphasize the importance in certain settings of the *ex post* wage-bargaining power of individual employees, and explore its effect on an employer's hiring decisions and choice of technology. They argue that where employees possess hold-up power arising from the specificity of their human capital, firms respond by hiring more workers to dilute their individual bargaining power, leading to allocative inefficiency. Specifically, profit maximization in this environment dictates that additional workers are hired until the bargaining-determined wage equals the worker's outside option (i.e., the competitive wage), so that the quasi-rents accruing to the relationship-specific human capital vanish. As a by-product of this allocative inefficiency, technical inefficiency may also arise to the extent that firms choose less productive technologies to offset reliance on over-employment as the sole means of reducing the bargaining power of workers. In short, the Stole-Zwiebel hypothesis predicts both allocative inefficiency in the form of over-employment and technical inefficiency to mitigate the effects on profitability of the resulting input distortions originating from the opportunistic behavior revealed in the intra-firm bargaining process.

Depken et al. (2001) present an alternative model that connects a firm's organizational design to its choice of an inefficient technology. In their model, an employee's power to withhold productive effort through shirking, once hired, leads firms to underutilize or hoard labor, and this is a source of intra-firm inefficiencies. Because it is costly to control shirking both directly (through monitoring, supervision, and teamwork) and indirectly (by hiring fewer employees, thereby increasing the marginal value of workers' effort), firms face a trade-off. To the extent that allocative inefficiency increases as the firm uses less direct methods to elicit work effort, technical inefficiency is deliberately introduced (by increasing labor set-up costs) to mitigate the allocative distortions. This model predicts that firms which use less direct means to reduce shirking exhibit greater labor hoarding. Therefore, the adoption by firms of allocatively and technically inefficient production methods in Depken et al. (2001) is a result of their use of labor hoarding to control shirking, rather than their over-employment of labor to eliminate quasi-rents arising from the possibility of employee hold up, as in Stole and Zweibel (1996). The predictions of their model were empirically tested and found to be consistent with the contrasting behavior of private and collective dairy farms in Jewish Palestine.

As Stole and Zweibel (1996a, p. 205) acknowledge, conducting an empirical test of the predictions of their model is complicated by two considerations. First, in the absence of detailed observations on the mechanism within the firm that determines wages, it is difficult to model the substitutability between capital and labor and this makes estimation of technology parameters problematic. Second, in their model the firm's induced production function includes the market wage as an argument. As a result, some of the usual properties of the associated cost function (e.g., linear homogeneity in factor prices) do not hold. Therefore, the standard econometric framework developed from duality theory for estimating technical and allocative inefficiency in a system of profit, production, and/or cost-share equations is inappropriate. We circumvent these difficulties by adopting a shadow-price approach and imposing an assumption that technology is characterized by decreasing returns to scale in the variable inputs.

We test two key predictions of the Stole-Zwiebel model. First, their model predicts that firms over-employ labor as a means of diluting the hold-up power of incumbent employees stemming from the wage-bargaining process. Second, the distortions introduced through 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, we first test for the presence of allocative inefficiency. This is tantamount to determining whether or not neoclassical profit maximization holds. It is important to note that rejection of neoclassical profit maximization is only a necessary condition for the Stole-Zwiebel hypothesis to be consistent with the data. When this necessary condition is met, the second step of our procedure is carried out to determine whether the front-load factor is equal to one, as predicted by the Stole-Zwiebel model. The profit of a Stole-Zwiebel firm with n employees is equal to the average of the neoclassical profit for firms of all sizes up to n employees. We use predicted values for neoclassical profit to construct an estimate of each firm's front-load factor. We then test the joint null hypothesis that the front-load factor for all firms in the sample is equal to one.

We implement this two-step procedure 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 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-Zweibel 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 fewer profits than his or her counterpart at a commercial bank. Loan officers at savings banks have a comparative advantage in "relationship banking," and their personal knowledge of the credit-worthiness of potential borrowers, often based on local factors that are unobservable to commercial-bank competitors, is a valuable asset to their employer. Once employed, these loan officers are in a position to expropriate any quasi-rents arising from the customer-specific information they possess by bargaining for a higher wage or threatening to quit and take this "knowledge capital" with them. According to the Stole-Zweibel model, savings banks are predicted to anticipate the potential hold-up power possessed by its specialized employees and hire more of them than is dictated by the strict application of neoclassical profit maximization. As a result, savings banks are predicted to hire an allocatively inefficient number of workers, seemingly sacrificing profits in the process.

Our application of the Stole-Zweibel model to a situation where mobile "knowledge workers," once employed, are in a position to hold up the firm is similar to the set up of Rebitzer and Taylor (2007), who examine the consequences of client-specific knowledge for the organizational structure of law firms, and the model of Anand and Galetovic (2006), who explore the implications of client-specific relationships for the market structure in investment banking. Rebitzer and Taylor (2007) 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 to the winners partner status which conveys an equity stake that becomes worthless if the partner leaves the firm. Thus, the firm in this model uses the employment contract, specifically controlling the acquisition of client-specific capital and the residual claim to profits, rather than the number of workers hired to solve the hold-up problem. Anand and Galetovic (2006) advance a complementary, market-based resolution of the hold-up problem, examining the roles of different dimensions of competition in maintaining incentives to develop relationship-specific capital in investment banking. By identifying adjustments in market equilibrium that facilitate these relationships, their analysis rationalizes certain empirical regularities in the oligopolistic market structure of the investment banking industry.

To preview our empirical results, we first provide evidence of allocative inefficiency, 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. These findings suggest that the firm-based resolution of the hold-up problem proposed by Stole and Zweibel, wherein employers adjust their hiring behavior to the wage-bargaining environment, may also play a role in maintaining incentives to develop relationship-specific capital in the banking sector.

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

II. The Stole-Zwiebel Hypothesis

Stole and Zwiebel (1996a) view the intra-firm bargaining process, which is absent in the neoclassical theory of the firm, as essential to understanding the firm's behavior. Intra-firm bargaining, which determines employees' wages and the firm's profit, is predicated on the assumption that workers and the firm split the joint surplus, which is 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

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

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

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

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

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 once the labor contract is signed and production begins. The firm chooses levels of labor n and other inputs 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

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

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

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

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

indicating that the wage-bargaining firm employs labor and other inputs up to the point where its profit is equal to the profit that would be earned in the competitive (neoclassical) case 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.

[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.

III. Testing the Stole-Zwiebel Hypothesis

The induced production function (4) implied by the Stole-Zwiebel hypothesis does not posses 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 allocative inefficiency in the use of variable inputs, and the second 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:

Model A	Neoclassical profit maximization holds or it fails to hold due to technical inefficiency.
Model B	Neoclassical profit maximization fails to hold due to allocative inefficiency or to both allocative and technical inefficiency.

If Model A is correct, then there is no allocative inefficiency and the wage-determination process generating the data does not distort input choices. If Model B 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 which model better explains the data.

Model A

Model A specifies the firm's profit function 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 convention, the profit function can be redefined 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 ε, is added to (8a) in estimation. By applying the Shephard-Uzawa-McFadden lemma to (8a), we obtain the cost-share equations

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

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

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

Model B

Model B incorporates the additional possibility that the firm fails to achieve allocative efficiency. To implement a duality-based estimation procedure, 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 shadow input prices $w^S = \theta w$, where θ is a vector that distorts the market prices w, leading to the first-order conditions

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

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.

The shadow profit function is defined by

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

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

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

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. Thus, we estimate

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

together with the actual cost-share equations

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

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

Our empirical estimates of Models A and B, reported in Table 3 and discussed below, reveal that Model B better explains the data and therefore provide evidence of allocative inefficiency. Hence, we proceed to step two and estimate the front-load factors for the firms in our sample. Since the front-load factor is the ratio of Stole-Zweibel profit to neoclassical profit, and the former is based on the sum of neoclassical profits earned over all inframarginal employment levels, we use Model B to obtain predicted neoclassical profits for alternative levels of employment.

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

$$\ln \pi_j = \beta_1 emp_j + \beta_2 emp_j^2 + \beta_3 TE_j + \psi Z_j + \varepsilon_j , \qquad (14)$$

where π_j denotes the predicted neoclassical profit of firm *j* 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 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 calculate the Stole-Zweibel profit for a firm with *n* employees. We then calculate each firm's front-load factor using equation (6b). Using the equivalence of $\gamma(F,n)=1$ to $\pi(n)=\tilde{\pi}(n)$, the null hypothesis for an individual firm *j* with *n* employees is

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

A distributional assumption about the estimated front-load factor is crucial to this test. Without such an assumption, the estimated covariance or correlation could not provide a measure of how close the two series, $\ln \pi(n)$ and $\ln \tilde{\pi}(n)$, are. Accordingly, we assume that the estimated front-load factor follows a normal distribution. This assumption follows from the assumption that the error term in equation (14) is normally distributed. Therefore, $\ln \hat{\pi}_j(n)$ 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-Zweibel profit to neoclassical profit (or the front-load factor) of an individual firm equals one. However, there are insufficient data in the sample to carry out such a test at the level of the firm. Consequently, our test of the null hypothesis that the front-load factor equals one is performed at the industry level:

$$H_O : \delta_1 = 0, \ \cdots, \ \delta_j = 0, \ \cdots, \ \delta_N = 0 \qquad (16)$$

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

$$\hat{F} = \hat{\delta}^T \hat{\Sigma}^{-1} \hat{\delta} \qquad (17)$$

follows an F distribution with $N-1$ and $N-K-1$ degrees of freedom, where $\hat{\delta}^T = [\hat{\delta}_1, \cdots, \hat{\delta}_N]$. A derivation of the distribution of the test statistic is provided in Appendix 2.

IV. 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 the working class and small-and-medium-sized enterprises, and began operation in 1972. These savings banks evolved to compete with commercial banks for deposits and loans after the financial deregulation of the late 1980s and early 1990s. However, payment settlements and foreign-currency services are performed only by commercial banks.

The banking industry possesses two unique characteristics that facilitate a test of the Stole-Zwiebel hypothesis. First, banks are subject to regulatory supervision which 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 (2003) and introduce the quasi-fixed inputs they suggest to control for the 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;[2] 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 operating revenue divided by total assets net of physical 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

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

and 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 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

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

are estimating a translog profit function, our sample is limited to the 42 banks reporting positive profit every year.[3] 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)][4]. With fixed-effects estimation, allocative inefficiency can be parameterized as Atkinson and Cornwell (1994) suggested in their estimation of a cost system. However, it is extremely difficult computationally to separate firm-specific unobserved effects 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 propose to avoid this computational difficulty by imposing the assumption that the production process is homogeneous of degree $r<1$ in the variable inputs. We

[3] If a single equation were being estimated, then the rescaling method (making negative profits positive by adding a constant number to the negative profits) or the indicator method (making negative profits equal to 1 by adding an indicator variable to the left-hand size) could be utilized (Bos and Koetter, 2009). However, neither method is appropriate for 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.

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

exploit this assumption in our estimation procedure by taking advantage of the following definition and propositions.[5]

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: $\hat{\pi}(tw,z) = t^{\frac{r}{r-1}}\hat{\pi}(w,z)$ if and only if $F(v,z)$ is homogeneous of degree *r* in *v*.

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*.

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 as a factor multiplying the normalized profit function. Proposition 3 asserts that the linear homogeneity property of the profit function is

[5] Propositions 1 and 2 were established by Lau (1978) in Theorems II-1 and III-1, respectively. For completeness, we provide brief proofs of these propositions in Appendix 3.

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

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

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

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

with the symmetry restrictions, $\alpha_{kj} = \alpha_{jk}$ and $\beta_{ml} = \beta_{lm}$ imposed. 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

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

When quasi-fixed inputs are included, the system of equations given in (8a) and (8b) is given by (19) 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

$$u_{it} = u_{t0} + u_{t1} t + u_{t2} t^2, \qquad (21)$$

where u_{t0} is a bank-specific parameter, but parameters u_{t1} and u_{t2} are common to all banks. It is possible to specify more flexible forms of technical inefficiency.[6] However, since N = 42 is large and T = 7 is small, we use this more parsimonious specification. 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 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 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.[7]

V. Empirical Results

The results of estimating the system of equations in (19) and (13b) for Model A (where $\theta_k = 1\ \forall k$) and for Model B (where θ_k is unrestricted) are reported in Table 3. It is

---

[6] For example, as in Atkinson and Primont (2002), technical inefficiency could be specified as $u_{it} = u_{i0} + u_{i1}t + u_{i2}t^2$ where u_{i0}, u_{i1}, and u_{i2} are firm-specific parameters. However, this specification requires the estimation of 3N parameters, where N is the number of firms.

[7] 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 inefficiency (u_i) is affected by the exogenous variable because inefficiency (u_i) is derived from a distributional assumption such as half-normal or truncated normal. However, this problem does not arise in our model because we use the fixed-effects estimator for which a distributional assumption on inefficiency (u_i) is not required. For details, see Greene (2008, pp. 155-156).

expected that the presence of allocative inefficiencies will affect the estimates of the coefficients, particularly 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 in the two models, and the allocative inefficiency for borrowed money is smaller than for labor. However, the magnitude of this allocative inefficiency is small, which is consistent 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 finding 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) since the theoretical value for α_0 is greater than one.

Using these estimates, a test of the null hypothesis

$$H_0 : \theta_1 = \theta_2 = 1 \quad (22)$$

(that allocative inefficiency is absent) is performed. This test allows us to determine whether the unrestricted Model B or the restricted Model A better explains the data. The log-likelihood ratio

$$\hat{LR} = 2(LogL_B - LogL_A) \quad (23)$$

follows a χ^2 distribution with two degrees of freedom (equal to the number of restrictions imposed in Model A), where $LogL_A$ is the maximum value of the likelihood function for Model A and $LogL_B$ is the maximum value of the likelihood function for Model B. 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. Moreover, the estimated shadow wage rate is 51 percent of the actual wage rate, which is consistent with the Stole-Zwiebel prediction of overemployment. As shown in Figure 2, Korean savings banks are also technically inefficient. Despite moderate improvement over time, the magnitude of this technical inefficiency is large.[8] These allocative and technical inefficiencies may explain why savings banks in Korea are less profitable than commercial banks.

Given our finding of allocative inefficiency, we estimate the relationship between neoclassical profit and the level of employment specified in (14). Along with a firm's time-mean front-load factor, we perform between estimation. The estimated regression is

$$\begin{aligned}\ln \hat{\pi}_{j.} = &\underset{(0.196)}{6.438} \underset{(0.0061)}{+0.011}\, emp_{j.} \underset{(0.000027)}{-0.000027}\, emp_{j.}^2 \underset{(0.3965)}{+1.4052}\, TE_{j.} \\ &\underset{(0.0089)}{+0.0105}\, z_{1,j.} \underset{(0.0028)}{+0.0208}\, z_{2,j.} \underset{(0.0068)}{-0.0032}\, z_{3,j.} \qquad (24) \\ &\bar{R}^2 = 0.845\end{aligned}$$

where robust standard errors are in parentheses. These results identify a (weakly) quadratic relation between neoclassical profit and the level of employment, controlling for the choice of technology.[9] This concave relation is consistent with the predictions of the Stole-Zweibel model. Using this empirical relation between neoclassical profit and the level of employment, each firm's front-load factor is constructed using the current

[8] Berger and Humphrey (1997) concluded from their survey of the relevant literature that mean technical inefficiency for U.S. banks is 16%.
[9] In addition, fixed-effects and random-effects estimates of (24), which are not reported, capture this quadratic relation.

level of employment and the industry-wide average level of technical efficiency. The results are shown in Figure 3.

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 is 1.73. Thus, the null hypothesis that all of the front-load factors are unity cannot be rejected.

VI. Conclusions

We develop and implement a procedure for testing two key predictions of the Stole-Zwiebel model of intra-firm wage bargaining. First, 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 since firms are willing to sacrifice productive efficiency if that will enhance their profitability, given the wage-bargaining process. Second, 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.

The first step in our approach tests for the presence of allocative inefficiency. If there is no evidence of allocative inefficiency, then we reject the Stole-Zwiebel hypothesis. To allow for the estimation of both allocative and technical inefficiency in a panel of firms, we adopt the shadow-price approach and impose the maintained hypothesis that the production technology is homogeneous of degree $r < 1$ in the variable inputs. Using panel data on Korean savings banks, we then estimate a translog system of profit and cost-share equations.

Our first set of empirical results provides evidence of allocative and technical inefficiency that is consistent with over-employment in Korean savings banks. Therefore, we proceed to the second step and determine each firm'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.

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

$$Y = XB + \varepsilon \qquad \text{(A1-1)}$$

where $Y = \log\pi$, $\mathrm{X} = [emp, emp^2, TE, z_1, z_2, z_3]$ 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

$$y_{j.} = x_{j.}{}^T\beta + \alpha \ + \varepsilon_{j.} \qquad \text{(A1-2)}$$

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

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

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

$$E\left(y_{j.}(n_{j.};TE_{j.},Z_{j.})\right) = \left(x_{j.}\right)^T \beta + \alpha$$

$$Var\left(y_{j.}(n_{j.};TE_{j.},Z_{j.})\right) = Var\left(\frac{\sum_{t=1}^{T} y_{jt}(n_{jt};TE_{jt},Z_{jt})}{T}\right) = \frac{\sigma_\varepsilon^2}{T}. \qquad \text{(A1-4)}$$

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

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

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

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

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

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

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

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

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

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

$$\frac{\hat{F}}{N-1} = \left(\frac{\hat{\delta}^T \Omega^{-1} \hat{\delta}}{\hat{\sigma}_\varepsilon^{\,2}} \right) / (N-1)$$

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

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,

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

Appendix 3

In this appendix we provide proofs of Propositions 1-3.

Proof of Proposition 1

The production function $F(v,z)$ is homogeneous of degree r in v if and only if $F(tv,z) = t^r F(v,z)$ for all t > 0. Then, the normalized variable profit function can be written as

$$\begin{aligned} \hat{\pi}(tw,z) &\equiv \max_v \left\{ F(v,z) - \sum_{k=1}^{K} \left(tw_k \right) v_k \right\} \\ &= t^{-r} \max_v \left\{ F(tv,z) - \sum_{k=1}^{K} \left(t^r w_k \right) \left(tv_k \right) \right\} \\ &= t^{-r} \hat{\pi}(t^r w, z). \end{aligned}$$

Hence, we have $t^r\hat{\pi}(tw,z)=\hat{\pi}(t^{r-1}tw,z)$ and, by letting $t^{r-1}=\kappa,$ so that $t=\kappa^{\frac{1}{r-1}}$, we arrive at $\kappa^{\frac{r}{r-1}}\hat{\pi}(tw,z)=\hat{\pi}(\kappa tw,z)$, showing that $\hat{\pi}(w,z)$ is homogeneous of degree $\dfrac{r}{r-1}$ in *w*.

Proof of Proposition 2

The variable profit function (18) is related to the normalized variable profit function by

$$\pi(pe^u,w,z)=pe^u\hat{\pi}\left(\frac{w}{pe^u},z\right).$$

Letting $t=\left(pe^u\right)^{-1}$, Proposition 1 allows us to write $\hat{\pi}(tw,z)=t^{\frac{r}{r-1}}\hat{\pi}(w,z)$ if and only if $F(v,z)$ is homogeneous of degree *r* in *v*. Thus, we have

$$\begin{aligned}\pi(pe^u,w,z)&=t^{-1}\hat{\pi}\left(tw,z\right)\\&=t^{\frac{1}{r-1}}\hat{\pi}\left(w,z\right)\\&=\left(pe^u\right)^{\frac{1}{1-r}}\hat{\pi}\left(w,z\right).\end{aligned}$$

Proof of Proposition 3

Linear homogeneity of the variable profit function (18) in the prices (pe^u,w) is expressed as $\pi(\kappa pe^u,\kappa w,z)=\kappa\pi(pe^u,w,z)$ for all $\kappa>0$. Proposition 2 allows us to rewrite the left-hand side as $\pi(\kappa pe^u,\kappa w,z)=\left(\kappa pe^u\right)^{\frac{1}{1-r}}\hat{\pi}(\kappa w,z)$ and the right-hand side as $\kappa\pi(pe^u,w,z)=\kappa\left(pe^u\right)^{\frac{1}{1-r}}\hat{\pi}(w,z)$ if and only if $F(v,z)$ is homogeneous of degree *r* in *v*. Thus, we have

$$\left(\kappa pe^{u}\right)^{\frac{1}{1-r}}\hat{\pi}(\kappa w,z)=\kappa\left(pe^{u}\right)^{\frac{1}{1-r}}\hat{\pi}(w,z)$$

$$\hat{\pi}(\kappa w,z)=\kappa^{\frac{r}{r-1}}\hat{\pi}(w,z).$$

Therefore, linear homogeneity of $\pi(pe^{u},w)$ in prices is equivalent to $\hat{\pi}(w,z)$ being homogeneous of degree $\frac{r}{r-1}$ in *w*.

References

Aigner, Dennis J., C. A. Knox Lovell and Peter Schmidt. 1977. "Formulation and Estimation of Stochastic Frontier Production Function Models." *Journal of Econometrics*, 6 (1), pp. 21-37.

Alvarez, Antonio, Christine Amsler, Luis Orea, and Peter Schmidt. 2006. "Interpreting and Testing the Scaling Property in Models Where Inefficiency Depends on Firm Characteristics." *Journal of Productivity Analysis*, 25 (3), pp. 201-212.

Anand, Bharat N. and Alexander Galetovic, 2006. "Relationships, Competition and the Structure of Investment Banking Markets." *Journal of Industrial Economics*, 54 (2), pp. 151-199.

Atkinson, Scott E. and Robert Halvorsen. 1980. "A Test of Relative and Absolute Price Efficiency in Regulated Utilities." *Review of Economics and Statistics*, 62 (1), pp. 81-88.

Atkinson, Scott E. and Christopher Cornwell. 1994. "Parametric Estimation of Technical and Allocative Inefficiency with Panel Data." *International Economic Review*, 35 (1), pp. 231-243.

Atkinson, Scott E. and Daniel Primont. 2002. "Stochastic Estimation of Firm Technology, Inefficiency, and Productivity Growth using Shadow Cost and Distance Functions." *Journal of Econometrics*, 108 (2), pp. 203-225.

Berg, Sigbjorn A., Finn R. Forsund, and Eilev S. Jansen. 1992. "Malmquist Indices of Productivity Growth during the Deregulation of Norwegian Banking." *Scandinavian Journal of Economics*, 94, pp. S211-S228.

Berger, Allen N. and David Humphrey. 1997. "Efficiency of Financial Institutions: International Survey and Directions for Future Research." *European Journal of Operational Research*, 98 (2), pp. 175-212.

Berger, Allen N. and Loretta J. Mester. 1997. "Inside the Black Box: What Explains Differences in the Efficiencies of Financial Institutions?" *Journal of Banking and Finance*, 21 (7), pp. 895-947.

Bos, Jaap W. B. and Michael Koetter. 2009. "Handling Losses in Translog Models." *Applied Economics*, forthcoming.

Coase, Ronald H. 1937. "The Nature of the Firm." *Economica,* 4 (16), pp. 386-405.

Cornwell, Christopher, Peter Schmidt, and Robin Sickles. 1990. "Production Frontiers with Cross-Sectional and Time-Series Variation in Efficiency Levels." *Journal of Econometrics*, 46 (1-2), pp. 185-200.

Depken, Craig A. II, Esther Redmount, and Arthur Snow. 2001. "Shirking and the Choice of Technology: A Theory of Production Inefficiency with an Empirical Application." *Journal of Economic Behavior and Organization,* 44 (4), pp. 383-402.

Hart, Oliver D. and John Moore. 1988. "Incomplete Contracts and Renegotiation." *Econometrica,* 56 (4), pp. 755-785.

Hart, Oliver D. and John Moore. 1990. "Property Rights and the Nature of the Firm." *Journal of Political Economy*, 98 (6), pp. 1119-1158.

Hughes, Joseph P. and Loretta J. Mester. 1993. "A Quality and Risk-Adjusted Cost Function for Banks: Evidence on the "Too-Big-To-Fail" Doctrine." *Journal of Productivity Analysis,* 4 (3), pp. 293-315.

Greene, William H. 2008. “The Econometric Approach to Efficiency Analysis” in *The Measurement of Productive Efficiency and Productivity Growth.* Harold O. Fried, C.A. Knox Lovell, and Shelton S. Schmidt, eds. Oxford: Oxford University Press, Chapter 2.

Grossman, Sanford J. and Oliver D. Hart. 1986. “The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration.” *Journal of Political Economy*, 94 (4), pp. 691-719.

Jensen, Michael C. and William H. Meckling. 1976. “Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure.” *Journal of Financial Economics,* 3 (4), pp. 305-360.

Joskow, Paul. 1987. “Contract Duration and Relationship-Specific Investments: Empirical Evidence from Coal Markets.” *American Economic Review,* 77 (1), pp. 168-185.

Klein, Benjamin, Robert G. Crawford, and Armen A. Alchian. 1978. “Vertical Integration, Appropriable Rents, and the Competitive Contracting Process.” *Journal of Law and Economics,* 21 (2), pp. 297-326.

Kumbhakar, Subal C. 2001. “Estimation of Profit Functions When Profit Is Not Maximum.” *American Journal of Agricultural Economics,* 83 (1), pp. 1-19.

Kumbhakar, Subal C. and Efthymios G. Tsionas. 2005. “Measuring Technical and Allocative Inefficiency in the Translog Cost System: A Bayesian Approach.” *Journal of Econometrics,* 126 (2), pp. 355-384.

Lau, Lawrence J. 1978. "Applications of Profit Functions," in *Production Economics: A Dual Approach to Theory and Applications*. Melvyn Fuss and Daniel L. McFadden, eds. Amsterdam: North-Holland, Chapter 3.

Lau, Lawrence J. and Pan A. Yotopoulos. 1971. "A Test for Relative Efficiency and Application to Indian Agriculture." *American Economic Review*, 61 (1), pp. 94-109.

Mester, Loretta J. 1996. "A Study of Bank Efficiency Taking into Account Risk-Preference." *Journal of Banking and Finance*, 20 (6), pp. 1025-1045.

Nickerson, Jack A. and Brian S. Silverman. 2003. "Why Aren't All Truck Drivers Owner-Operators? Asset Ownership and the Employment Relation in Interstate For-Hire Trucking." *Journal of Economics and Management Strategy*, 12 (1), pp. 91-118.

Rebitzer, James B. and Lowell J. Taylor. 2007. "When Knowledge Is an Asset: Explaining the Organizational Structure of Large Law Firms." *Journal of Labor Economics*, 25 (2), pp. 201-229.

Schmidt, Peter and Robin C. Sickles. 1984. "Production Frontiers and Panel Data." *Journal of Business and Economic Statistics*, 2 (4), pp. 367-374.

Sealey, Calvin W. and James T. Lindley. 1977. "Inputs, Outputs, and a Theory of Production and Cost at Depository Financial Institutions." *Journal of Finance*, 32 (4), pp. 1251-1266.

Stole, Lars A. and Jeffrey Zwiebel. 1996a. "Organizational Design and Technology Choice under Intrafirm Bargaining." *American Economic Review*, 86 (1), pp. 195-222.

Stole, Lars A. and Jeffrey Zwiebel. 1996b. "Intra-firm Bargaining under Non-binding Contracts." *Review of Economic Studies*, 63 (216), pp. 375-410.

Vukina, Tomislav and Porametr Leegomonchai. 2006. "Oligopsony Power, Asset Specificity, and Hold-up: Evidence from the Broiler Industry." *American Journal of Agricultural Economics*, 88 (3), pp. 589-605.

Wang, Hung-jen and Peter Schmidt. 2002. "One-Step and Two-Step Estimation of Exogenous Variables on Technical Efficiency Levels." *Journal of Productivity Analysis*, 18 (2), pp. 129-144.

Wheelock, David C. and Paul W. Wilson. 1999. "Technical Progress, Inefficiency, and Productivity Change in U.S. Banking, 1984-1993." *Journal of Money, Credit, and Banking*, 31 (2), pp. 212-234.

Whinston, Michael D. 2001. "Assessing the Property Rights and Transaction-Cost Theories of Firm Scope." *American Economic Review Papers and Proceedings*, 91 (2), pp. 184-188.

Whinston, Michael D. 2003. "On the Transaction Cost Determinants of Vertical Integration." *Journal of Law, Economics and Organization*, 19 (1), pp. 1-23.

Williamson, Oliver E. 1979. "Transaction-Cost Economics: The Governance of Contractual Relation." *Journal of Law and Economics*, 22 (2), pp. 233-261.

Williamson, Oliver E. 1985. *The Economic Institutions of Capitalism*. New York: Free Press.

Table 1: *Descriptive statistics of savings (commercial) banks*

	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	20.9	20.4	3.1	-53.4	84.2	87.3	43.7
(₩ million)	(35.9)	(38.3)	(13.2)	(70.0)	(105.0)	(98.2)	(107.0)
Wages per worker	28.7	33.9	37.8	40.1	45.7	49.0	48.5
(₩ million)	(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)

Table 2: *Means and standard deviations of variables*[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 percentage 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 off-balance sheet amount, financial capital, and physical capital, while π denotes variable profit.

Table 3: *Empirical Results*[1]

Variables	(parameters)	Model A		Model B	
$\ln p$	(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)			0.5137	(0.0730)
	(θ_2)			0.8283	(0.0492)

[1] Standard errors are in parentheses.

Figure 1: *Profit maximizing choice of labor*

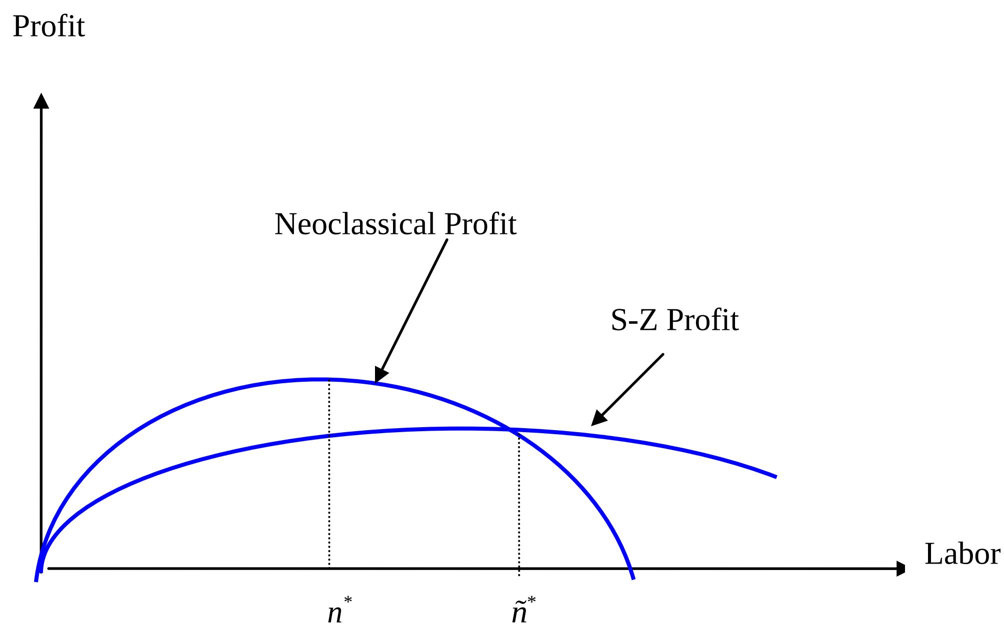

Figure 2: *Mean values and standard deviations of technical efficiency from Model A (in blue) and Model B (in red)*

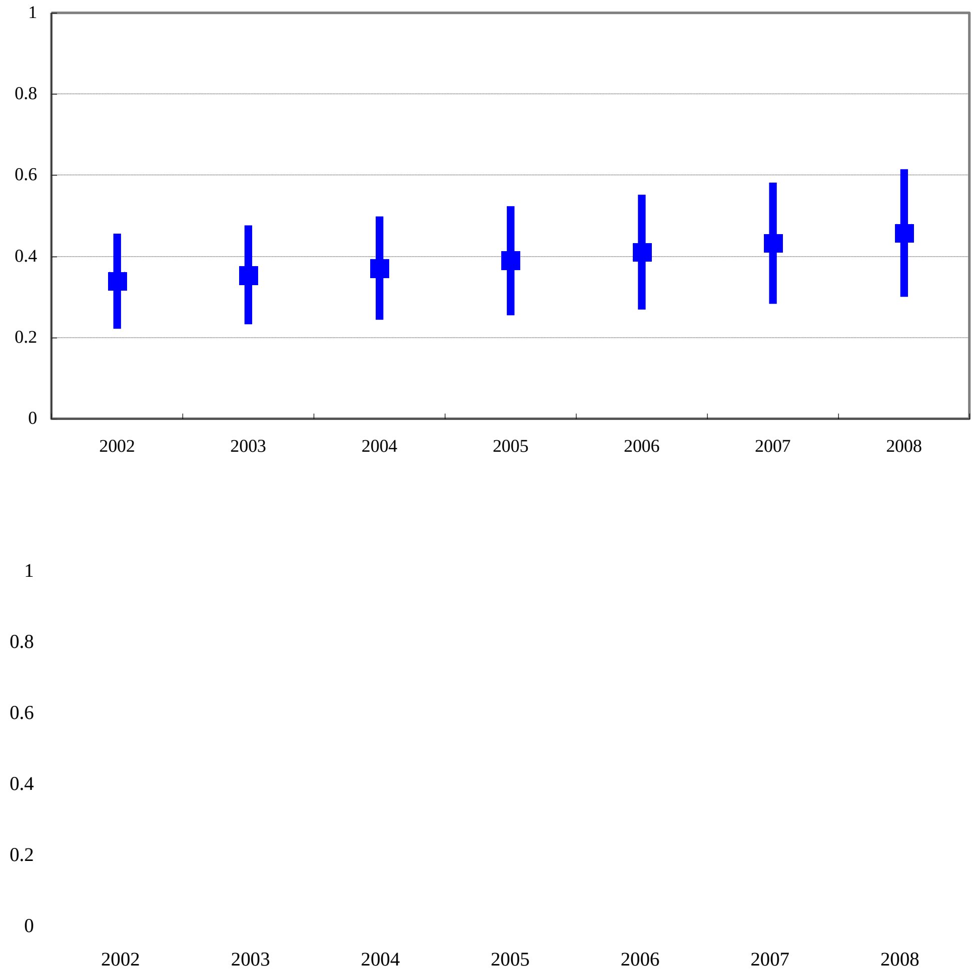

Figure 3: *Each firm's constructed front-load factor*