Theory of the firm and structure of residual rights

Xiaokai Yang *, Yew-Kwang Ng

Department of Economics, Monash University, Clayton, Victoria 3168, Australia

Received August 1991, final version received March 1993

Abstract

This paper develops a general equilibrium model with consumer–producers, economies of specialization, and transaction costs to investigate the emergence of firms from the division of labor and the function of a structure of residual rights. It is shown that the institution of the firm can be used to get intangible intellectual property involved in the division of labor while avoiding its direct pricing and marketing, so that the division of labor can be promoted by saving on transaction costs.

JEL classification: D21; D23

1. Introduction

This paper uses a formal version of the Coase–Cheung theory of the firm to investigate the productivity implications of the structure of residual rights. According to Coase (1937), the rationale for the existence of the firm is the differences in transaction costs between the market and the institution of the firm. If the ‘market’ involves greater transaction costs than the firm, then the division of labor will be organized within a firm rather than via the market. Hence, the boundary of the firm is determined by the equalization condition of marginal benefits (saving on

* The authors are grateful to Jeff Borland, Richard Day, Gene Grossman, Edwin Mills, Paul Milgrom, Yingyi Qian, M. Aoki, and an anonymous referee for their comments. Thanks also go to the Ford Foundation and the Open Society Fund for financial support. The second author wishes to acknowledge the predominant contribution of the first author.

* Corresponding author.

0167-2681/95/$09.50 © 1995 Elsevier Science B.V. All rights reserved

SSDI 0167-2681(94)00009-4
transaction costs relative to the market) and marginal bureaucracy costs of the firm.

For a further development, Cheung (1983) argued that the firm does not replace the market with a non-market institution. Rather, it replaces the market for intermediate goods with the market for labor hired to produce the intermediate goods. Hence, the firm emerges if transaction costs are higher in trading intermediate goods than in trading the labor hired to produce the intermediate goods. Thus, the argument that the firm internalizes externalities does not make sense since externalities in the market for goods may be replaced by externalities in the market for factors. ¹

Given the existence of the firm, it may have many different structures of ownership of a firm. Does the structure of ownership make a difference? According to Coase (1960), the structure of ownership matters if transaction costs are not negligible. We show here that the existence of transaction costs is necessary but not sufficient for the structure of ownership to have productivity implications. ²

Our story runs as follows. There are many ex ante identical consumer-producers in an economy. Each individual as a consumer must consume a final good, called cloth, the production of which requires an intermediate good, called management of production, as an input. Although each individual as a producer can produce either or both cloth and management service, he prefers specialized production due to economies of specialization. Each individual’s optimal decision is a corner solution. He may choose autarky which implies that he self-provides cloth and self-manages the production. This pattern of organization generates a low productivity level because economies of specialization cannot be exploited. Individuals can choose the division of labor which implies that some individuals become professional producers of cloth and others become professional producers of management service. This pattern of organization generates a high productivity of both goods, on the one hand, and incur transaction costs, on the other. Hence, there is a tradeoff between economies of specialization and transaction costs.

¹ The principal-agent theory and related contract theory formalize a tradeoff between efficient risk sharing and effective incentive mechanism to show that the second best partial equilibrium contract generates transaction costs (see Hart and Holmstrom, 1987). The procurement model (see, for example, Laffont and Tirole, 1986 and Lewis and Sappington, 1991) addresses the decision problem in choosing between outside procurement and self-provision within a firm by formalizing a tradeoff between exploitation of exogenous comparative advantage and distortions resulting from information asymmetry. Zhang (1993) identifies the condition under which capitalists act as principals. All the models do not explain why and how firms emerge from the division of labor.

² Grossman and Hart (1986) and Hart and Moore (1990), built on the work of Williamson (1975) and Klein, Crawford, and Alchian (1978), have established the statement that the structure of ownership makes a difference, although their theory is a theory of optimum ownership structure rather than a theory of the firm. However, their models established the statement on the basis of asset specificity argument. Our model will establish a similar statement on the basis of indirect pricing of intangible intellectual property rights. In the Holmstrom and Milgrom model (1994) and our model here the labor contract is emphasized as an essential feature of the institution of the firm.
If transaction efficiency is high, then the division of labor will occur in equilibrium because economies of specialization outweigh the transaction costs generated by the division of labor. Otherwise autarky will be chosen as the equilibrium. Suppose that transaction efficiency is so high that individuals prefer the division of labor to autarky. Then there are three different structures of residual rights which can be used to organize transactions required by the division of labor. The first (structure 1) is comprised of markets for cloth and management services. Specialist producers of cloth exchange cloth for management consultant service with specialist producers of management service. For this market structure, residual rights to any contracts and authority are symmetrically distributed between trade partners and no firms and labor market exist. The second structure of residual rights (structure 2) is comprised of the market for cloth and the market for labor hired to produce management service within a firm. The producer of cloth is the owner of the firm and specialist producers of management services are employees. Residual rights and authority are asymmetrically distributed between the employer and his employees. The employer claims the residual of the firm which is the difference between revenue and wage bill. The third structure of residual rights (structure 3) is comprised of the market for cloth and the market for labor hired to produce cloth within a firm. The professional manager is the owner of the firm and specialist producers of cloth are employees. For the final two structures of residual rights, the firm emerges from the division of labor. Compared with structure 1, the two structures involve a labor market but not a market for management services. As Cheung argued, the firm replaces the market for intermediate goods with the market for labor hired to produce the intermediate goods. Although both structures involve the firm and asymmetric structure of residual rights, they have different structures of ownership of a firm.

Assuming that transaction efficiency is much lower for management service than for labor, then the institution of the firm can be used to organize the division of labor more efficiently because it avoids trade in management services. Suppose further that transaction efficiency for labor hired to produce management services is much lower than for labor hired to produce cloth because it is prohibitively expensive to measure efforts exerted producing intangible management (a sort of intellectual property) and to measure output level (quality and quantity) of management services. Then the division of labor can be more efficiently organized in structure 3 than in structure 2 because structure 3 involves trade in cloth and in labor hired to produce cloth but not trade in management services and in labor hired to produce management services, while structure 2 involves trade in cloth and in labor hired to produce management services. Hence, structure 3 will occur in equilibrium if the transaction efficiencies for labor hired to produce cloth and for cloth are sufficiently high.

The claim to the residual of the firm by the manager is the indirect price of management services. Therefore, the function of the asymmetric structure of residual rights is to get the activity with the lowest transaction efficiency involved
in the division of labor while avoiding direct pricing and marketing of the activity, such that the division of labor and productivity are promoted. In a sense, the function of the asymmetric structure of residual rights is similar to that of a patent law which enforces rights to intangible intellectual property thereby promoting the division of labor in research and development. However, the asymmetric structure of residual rights can indirectly price those intangible intellectual properties which are prohibitively expensive to protect even through a patent law (can we enforce a patent law of intangible management knowledge?).

Although transaction costs play an important role in the analysis, we do not explicitly model the source of the transaction costs. Rather, 'iceberg'-type transaction costs are assumed. These transaction costs may be interpreted as anticipated costs of ex post strategic or opportunistic behavior associated with trade in a particular market. The exogeneity of transaction costs allows us to capture in a simple way the main ideas which seem to have emerged from the transaction cost literature, namely that transaction costs exist and that they may differ across goods and factors and across institutional structures for production and exchange. 3

After presenting our analysis, we use it for the concluding section to explain specific institutional settings, appertaining to Chinese history and to more recent Maxian issues in modern socialist economies.

2. A model with intermediate goods

Let us consider an economy with \( M \) identical consumer-producers. There is one consumer good and one intermediate good (or service) in this economy. The self-provided amounts of the consumer and intermediate goods are \( y \) and \( x \) respectively. The quantities of the two goods sold in the market are \( y^s \) and \( x^s \) respectively. The quantities of the two goods purchased in the market are \( y^d \) and \( x^d \) respectively.

In order to produce the final good, the intermediate good is a necessary input. An individual's production function for the final good is

\[
y + y^s = \left\{ \left[ x + (1 - t) x^d \right] L_y \right\}^a, \quad a < 1, \quad (1a)
\]

3 For example, Williamson (1985) argued that since relationships between economic agents are governed by incomplete contracts, there is scope for opportunistic behavior by the parties to such contracts, and that this possibility of opportunistic behavior constitutes a source of transaction costs. The scope for opportunistic behavior, and hence the size of transaction costs, will differ across contracts and across institutions which specify different rules for governing a relationship when unforeseen circumstances arise, and according to the frequency of trade and the degree of asset specificity and uncertainty involved in the particular relationship.

4 Equilibrium may not exist if \( a \geq 1 \).
where \( x^d \) is the amount of the intermediate good purchased from the market and \( t \) is its transaction cost coefficient. The fraction \( t \) of \( x^d \) disappears in transaction.\(^5\). Hence, \((1 - t)x^d\) is the amount an individual receives from the purchase of this intermediate good. The amount self-provided of this good is \( x \). \( L_y \) is the labor share in producing the consumer good and \( y + y^s \) is the output level of this consumer good. A person’s labor share in producing a good is defined as his level of specialization in producing this good.

A production function for a good is said to exhibit economies of specialization if total factor productivity of the good is an increasing function of the level of specialization in producing this good. Total factor productivity of good \( y \), \( (y + y^s)/(X^{a}L^{1-a}) = L_y^{2a-1} \) increases with the level of specialization \( L_y \) if \( a > 1/2 \), where \( X = x + (1 - t)x^d \) is the input level of the intermediate good. The parameter \( a \) represents the degree of economies of specialization in producing the final good.\(^6\)

The production function of the intermediate good is

\[
x + x^s + L_x^b, \tag{1b}
\]

where \( x + x^s \) is the output level of the intermediate good and \( L_x \) is a person’s level of specialization in producing the intermediate good. This production function displays economies of specialization if \( b > 1 \). The parameter \( b \) represents the degree of economies of specialization in producing the intermediate good.

It is assumed that labor is specific for an individual and there is a constraint on the total labor share, given by

\[
L_x + L_y = 1, \quad 0 \leq L_i \leq 1, \quad i = x, y. \tag{1c}
\]

By choosing a proper unit of labor, we can assume that each person is endowed with a unit of labor. \((1)\) specifies a system of production functions for an individual. According to these production functions, each individual may self-provide all goods.

Transaction technology of the final good is the same as that of the intermediate good. The fraction \( k \) of a shipment disappears in transit. Thus, \((1 - k)y^d\) is the amount an individual receives from the purchase of the consumer good. The amount consumed of the final good is \( y + (1 - k)y^d \). The utility function is identical for all individuals and given by

\[
U = y + (1 - k)y^d. \tag{2}
\]

Free entry for all individuals into any sector and a large \( M \) are assumed.

---

\(^5\) This specification of ‘iceberg’ transaction technology is necessary for avoiding notorious indexes of origins and destinations of deliveries.

\(^6\) Parameter \( a \) may be divided into two components \( a_1 \) and \( a_2 \), which represent output elasticities of the two inputs. However, this separation complicates the algebra and contributes little to the results.

\(^7\) In Yang (1990), the case of multiple final goods is tackled by adopting a Cobb–Douglas utility function, giving rise to a trade-off between preference for diverse consumption and economies of specialization.
3. Configuration and corner solution versus market structure and corner equilibrium

The set of feasible market structures can now be characterized. Each individual makes a decision about which goods to produce and on his demand for and supply of any traded good to maximize. A given structure of production and trade activities for any individual is defined as a configuration. There are $2^6 = 64$ combinations of zero and non-zero values of $x, x^s, x^d, y, y^s, y^d$ and therefore 64 possible configurations. The combination of configurations of the $M$ individuals in the economy is defined as a market structure, or simply a structure. A feasible market structure consists of a set of choices of configurations by individuals such that for any traded good, demand for the good is matched by supply of the good. There exists a corner equilibrium for each structure. Corner equilibrium is defined as a set of relative prices and relative numbers of individuals choosing different configurations such that (i) for any traded good market supply equals market demand (i.e. market clearing); and (ii) each individual maximizes utility at the given prices and for a given structure. With the assumption of free entry, (ii) implies that the utility of all individuals is equalized in any corner equilibrium.

Using the Kuhn–Tucker theorem it is possible to rule out the interior solution and many corner solutions from the list of candidates for the optimum decision.

Proposition 1. If $b > 1$ and $a \in (0.5,1)$, an individual sells at most one good and does not buy and self-provide the same good; he self-provides the consumer good when he sells it; and he does not self-provide the intermediate good unless he self-provides the final good.

The proof is given in the Appendix. In this paper, it is assumed that $b > 1$ and $a \in (0.5,1)$. Taking into account this proposition and the possibility for setting up firms, there are four structures. Here, institution of the firm is defined as a structure of transactions that satisfies the following conditions. (a) One party (employer) has control rights of other party’s (employee’s) labor; (b) In relevant contracts, only payment to employees is specified. An employer claims residuals of returns; and (c) An employer sells goods or services, produced using employees’ labor, in good market. This section first solves for corner equilibria in these four structures, then identifies the general equilibrium from these corner equilibria.  

---

* Note that the specific model in this paper can be used to solve for the explicit equilibrium, so that a general proof on the existence of equilibrium is not necessary. On the other hand, the major concern of this paper is the evolution of the division of labor and institutional arrangements. A specific model is necessary for comparative statics that describes such an evolution because the comparative statics generate shifts of general equilibrium across market structures and it can be characterized only by comparing corner equilibrium utilities in different structures. This is another restriction imposed by the complexity of non-linear programming technique.
Structure A (autarky)

Each individual self-provides all goods, so \( x^s = y^s = x^d = y^d = 0 \). Assuming this in the decision problem (1)-(2) yields

\[
\text{Max } U = y = x^a L^a_y = L^a_x (1 - L_x)^a,
\]

which implies

\[
L_x = b / (1 + b), \quad L_y = 1 / (1 + b), \quad x = [b / (1 + b)]^b, \quad \text{(3b)}
\]

\[
U_A = y = \left[ b^b / (1 + b)^{1+b} \right]^a.
\]

Here the individual maximum utility \( U_A \) is per capita real income as well as the maximum per capita output level of the final good in autarky.

Structure D

Individuals specialize in either the intermediate or final good. Denoting a configuration in which an individual sells the intermediate good and buys the final good by \((x/y)\) and a configuration in which an individual sells the final good and buys the intermediate good by \((y/x)\), structure D consists of configurations \((x/y)\) and \((y/x)\). Authority and residual rights are equally distributed between the two parties to a contract. There are two steps in solving for the corner equilibrium. The individual optimum decisions for these two configurations are first solved. Then individual demand and supply, generated by the optimum decisions, and the market clearing conditions, and utility equalization condition are used to solve for the corner equilibrium.

Let \( x^s, y^d > 0, L_x = 1, x = x^d = y = y^s = L_y = 0; (1b, c) \) and (2) give the individual decision for configuration \((x/y)\) as

\[
\begin{align*}
py^d &= x^s, \quad \text{(budget constraint)} \\
x^d &= L_x^a, \quad L_x = 1, \quad \text{(production function and endowment constraint)} \\
U_x &= y^d(1 - k) = L_x^a(1 - k)/p = (1 - k)p, \quad \text{(indirect utility function)}
\end{align*}
\]

where \( p \) is the price of the final good in terms of the intermediate good and \( U_x \) is the indirect utility function for configuration \((x/y)\).

Let \( y^s, x, y^d > 0, L_y = 1, y^d = x = x^s = L_x = 0 \) in \((1a, c)\) and (2); the individual decision problem for configuration \((y/x)\) is

\[
\begin{align*}
\text{Max } U_y &= y = \left[ (1 - t) x^d \right]^- a - x^d / p, \quad \text{(5a)} \\
\text{s.t. } y + y^s &= \left[ (1 - t) x^d L_y \right]^- a, \quad L_y = 1, \quad \text{(production function)} \\
py^s &= x^d, \quad \text{(budget constraint)}
\end{align*}
\]

where \( t \) is the transaction cost coefficient for the intermediate good. The optimum decisions are

\[
\begin{align*}
L_y &= 1, \quad x^d = \left[ ap(1 - t) \right]^{1/(1-a)}, \quad y^s = x^d / p, \quad \text{(5b)} \\
U_y &= y = (1 - a) \left[ ap(1 - t) \right]^{a/(1-a)}
\end{align*}
\]
where $U_y$ is the indirect utility function for configuration $(y/x)$.

Let $M_x$ and $M_y$ be the numbers of individuals selling the intermediate and final goods respectively. Multiplying $M_i$ $(i = x, y)$ with individual demand or supply gives market demand or supply. Equalizing the market demand to the market supply establishes the market clearing conditions. The assumption of free entry combined with the individuals' behaviour of maximizing utility ensures the utility equalization condition. From the utility equalization condition and market clearing condition, we can thus solve for the corner equilibrium relative number of individuals choosing two configurations, $M_{xy} = M_x/M_y$, and relative price $p$. The values of $M_x$ and $M_y$ are determined by $M_{xy}$ and the population size $M = M_x + M_y$.

The utility equalization condition is

$$U_y = (1 - a)\left[(1 - t)ap\right]^{a/(1-a)} = (1 - k)/p \equiv U_x.$$ (6)

(6) gives the corner equilibrium $p$ and the percapita real income in structure $D$, $U_D$:

$$p = \left[(1 - k)/(1 - a)\right]^{1-a} / \left[a(1 - t)\right]^a,$$ (7)

$$U_D = a^a(1 - a)^{1-a}(1 - t)^a(1 - k)^a.$$ (8)

The market clearing condition for the intermediate good is

$$M_x x^s = M_y x^d.$$ (9)

Note that the market clearing condition for the final good is not independent of (8) due to Walras' law. Inserting $x^d$ and $x^s$, given by the individual optimum decisions (4) and (5), and $p$, given by (7), into (8), we have

$$M_{xy} = x^d/x^s = x^d = a(1 - k)/(1 - a).$$ (9)

This $M_{xy}$ is the corner equilibrium relative number of individuals selling different goods in structure $D$.

**Structure FY**

Let $(y/L_x)$ denote a configuration in which an individual sells the final good, buys labor, produces the final good, and employs workers to produce the intermediate good. Similarly, $(L_x/y)$ denotes a configuration in which an individual sells his labor, buys the final good, and becomes a worker producing the intermediate good. Structure FY consists of configuration $(y/L_x)$ and configuration $(L_x/y)$. In
this structure, the producer of \( y \) claims the residual rights in the relationship. The individual decision problem for configuration \((y/L_x)\) is

\[
\begin{align*}
\text{Max} & \quad U_y = y = [(1 - v)N]^a - N/q, \\
\text{s.t.} & \quad y + y^* = [(1 - v)L_x^bNL_y]^{-1}, \quad L_x = 1, \quad L_y = 1, \\
& \quad (\text{production function and endowment constraint}) \\
& \quad qy^* = NL_x = N, \quad (\text{budget constraint or trade balance})
\end{align*}
\]

where \( v \) is the transaction cost coefficient of labor hired to produce the intermediate good, \( N \) is the number of workers hired by an employer, \( q \) is the price of the final good in terms of labor. \( L_y \) is the level of specialization of an individual choosing configuration \((y/L_x)\) in producing the final good and \( L_x \) is the level of specialization of an individual choosing configuration \((L_x/y)\) in producing the intermediate good. \( L_x \) and \( L_y \) are the employer’s decision variables because of asymmetric distribution of control rights.

The optimum decisions are

\[
L_x = L_y = 1, \quad y^* = N/q, \quad U_y = U_y = [(1 - v)N]^a - N/q, \quad (10b)
\]

\[
N = [(1 - v)^a aq]^{1/(1-a)}, \quad (10c)
\]

where \( U_y \) is the indirect utility of an employer.

For an individual choosing configuration \((L_x/y)\), all variables are fixed, given by

\[
U_{L_x} = (1 - k) y^d = (1 - k)/q, \quad (11a)
\]

\[
qy^d = L_x = 1, \quad (\text{budget and endowment constraints})
\]

where \( U_{L_x} \) is the indirect utility of an individual choosing configuration \((L_x/y)\). Here, an individual has one unit of labor and labor is the numeraire. \(^9\)

Manipulating the market clearing condition and utility equalization condition yields the corner equilibrium in structure FY, given by

\[
N = M_{xy} = (1 - k)a/(1 - a), \quad q = [(1 - k)/(1 - a)]^{1-a}/[a(1 - v)]^a, \quad (11b)
\]

\[
U_{FY} = [(1 - v)(1 - k)a]^{a}(1 - a)^{1-a},
\]

where \( U_{FY} \) is the percapita real income in structure FY, \( M_{xy} = M_x/M_y \) is the number of individuals choosing configuration \((L_x/y)\) relative to those choosing \((y/L_x)\), \( M_x \) is the total number of employees, and \( M_y \) is the total number of employees.

\(^9\) Note that for convenience in manipulating the algebra, the intermediate good is assumed to be the numeraire in market D, and labor the numeraire in structures FY and FX.
employers. It should be noted that an employer cannot manipulate $M_{xy}$ although he can choose $N$ (the number of workers hired by an employer) and that the corner equilibrium value of $N$ is the same as the corner equilibrium value of $M_{xy}$.\(^\text{10}\) While more than one worker may be employed, economies of specialization do not extend beyond a specific individual.

**Structure FX**

Let $(y/L_y)$ denote a configuration in which an individual sells the final good, buys labor, produces the intermediate good, and hires workers to produce the final good. Similarly, $(L_y/y)$ denotes a configuration in which an individual sells his labor, buys the final good, and becomes a worker employed producing the final good. Structure FX consists of configurations $(y/L_y)$ and $(L_y/y)$. The difference between structures FY and FX is that in structure FY a producer of the final good is the owner of the firm, whereas in structure FX a producer of the intermediate good is the owner of the firm.

Repeating the procedure of solving for the corner equilibrium in structure FY, the corner equilibrium in structure FX is given by

$$M_{yx} = (1 - k)(1 - a)/a, \quad q = \left[(1 - k)/a(1 - r)\right]^a/(1 - a)^{1-a}, \quad (12)$$

$$U_{Fx} = \left[(1 - r)a\right]^a \left[(1 - k)(1 - a)\right]^{1-a},$$

where $r$ is the transaction cost coefficient of labor hired to produce the final good in structure FX, $U_{Fx}$ is the per capita real income in structure FX, and $q$ is the price of the final good in terms of labor. $M_{yx}$ is the number of individuals producing the final good and choosing configuration $(L_y/y)$ relative to those producing the intermediate good and choosing configuration $(y/L_y)$. A comparison between the per capita real income in structures A, D, FY, and FX leads us to

**Proposition 2.**

1. Structure A generates the maximum per capita real income if transaction efficiency and economies of specialization are sufficiently small.
2. The structure with the division of labor and without firms generates the maximum per capita real income if transaction efficiency and economies of specialization are sufficiently large and the transaction efficiency for the intermediate good is higher than for labor.
3. The structure with the division of labor and with producers of the final good as the bosses of firms generates the maximum per capita real income if transaction efficiency and economies of specialization are sufficiently large and the transaction efficiency for the labor used to produce the intermediate good is great

\(^{10}\) To mistakenly believe that a single employer can choose $M_{xy}$ since he can choose $N$ is to commit the fallacy of attribution discussed by Ng (1982).
compared to that for the intermediate good and for the labor used to produce the final good.

(4) The structure with the division of labor and with producers of the intermediate good as the bosses of firms generates the maximum per capita real income if transaction efficiency and economies of specialization are sufficiently large and the transaction efficiency for the labor used to produce the final good is great compared to that for the intermediate good and for the labor used to produce the intermediate good. 11

The corner equilibrium with the maximum per capita real income is the general equilibrium because all corner equilibria satisfy all conditions for equilibrium, except that individuals' utilities are not maximized with respect to the choice of configurations across structures. 12 Thus, proposition 2 implies.

Corollary 1.

(1) The general equilibrium is the corner equilibrium in autarky if transaction efficiency and economies of specialization are sufficiently small.

(2) The general equilibrium is the corner equilibrium in the structure with the division of labor and without firms if transaction efficiency and economies of specialization are sufficiently large and the transaction efficiency for the intermediate good is higher than for labor.

(3) The general equilibrium is the corner equilibrium in the structure with the division of labor and with producers of the final good as the bosses of firms if transaction efficiency and economies of specialization are sufficiently large and the transaction efficiency for the labor used to produce the intermediate good is great compared to that for the intermediate good and for the labor used to produce the final good.

(4) The general equilibrium is the corner equilibrium in the structure with the division of labor and with producers of the intermediate good as the bosses of firms if transaction efficiency and economies of specialization are sufficiently large and the transaction efficiency for the labor used to produce the final good is great compared to that for the intermediate good and for the labor used to produce the intermediate good.

11 More technically, proposition 2 states: (1) Structure A generates the maximum per capita real income if \( f(k, t, a) < g(b) \), where \( f(k, t, a) = (1 - k)(1 - t)a(1 - a)^{1/a - 1} \), \( \delta f/\delta k < 0 \), \( \delta f/\delta t < 0 \), \( \delta f/\delta a > 0 \), \( g(b) = b^b/(1 + b)^{1 + b} \), \( \delta g/\delta b < 0 \). (2) Structure D generates the maximum per capita real income if \( f(k, t, a) > g(b) \), \( t < u \) and \( h(t, r, k, a) > 1 \), where \( h(t, r, k, a) = [(1 - t)/(1 - r)]^{(1 - k)^2a - 1} \), \( \delta h/\delta k < 0 \), \( \delta h/\delta t < 0 \), \( \delta h/\delta a > 0 \). (3) Structure FY generates the maximum per capita real income if \( f(k, v, a) > g(b) \), \( t > v \) and \( h(v, r, k, a) > 1 \), where \( \delta f/\delta v < 0 \), \( \delta h/\delta v < 0 \). (4) Structure FX generates the maximum per capita real income if \( f(k, r, a) > g(b) \), \( h(t, r, k, a) < 1 \), and \( h(v, r, k, a) < 1 \), where \( \delta f/\delta r < 0 \).

12 Indeed, Yang (1988a, 1990) has proved that all corner equilibria that do not generate the maximum per capita real income are not general equilibrium.
Corollary 1 implies that an economy will evolve from autarky to the division of labor as transaction efficiency is improved. The equilibrium institution based on the division of labor is comprised of the markets for the final and intermediate goods if transaction efficiency of the intermediate good is higher than that of labor. It is comprised of the markets for the final good and for the labor hired by firms to produce the final good if transaction efficiency of labor hired to produce the final good is higher than that hired to produce the intermediate good. It is comprised of the markets for the final good and for the labor hired by firms to produce the intermediate good if transaction efficiency of labor hired to produce the intermediate good is higher than that hired to produce the final good. Firms will emerge from the latter two institutional arrangements. Fig. 1 illustrates market structures A (autarky), D (division of labor without firms), FY (division of labor with firms owned by producers of the final good), and FX (division of labor with firms owned by producers of the intermediate good).

4. Structure of residual rights, economies of specialization, economies of division of labor, and economies of the firm

This section investigates first the relationship between economies of specialization, economies of division of labor, and economies of the firm, and then the implications of a structure of residual rights for the theory of the firm.

As shown in (1a), the production function for good y displays economies of specialization if \( a > 1/2 \) since the total factor productivity of good y increases with a person’s level of specialization in producing good y. Also, (1b) implies that the production function for good x exhibits economies of specialization in producing the intermediate good if \( b > 1 \) since labor productivity of good x increases with a person’s level of specialization in producing this good.
Economy of division of labor is a concept different from economy of specialization and economy of scale. There is a need for defining the concept of economy of division of labor before examining the difference between these concepts. The system of production function (1) is said to exhibit economies of division of labor if the maximum output level of the final good with zero transaction cost is greater in structure D than in structure A. In order to ascertain if production function (1) exhibits economies of division of labor, it is necessary to solve for the maximum output level of final good per capita in structure D. Since there is only one final good (and therefore utility is determined by per capita output level of the final good), the Pareto-efficient allocation in structure D will generate the maximum output level of the final good per capita. Maximizing utility for a configuration subject to the balance constraint between aggregate consumption and production and to the constraint that utility in another configuration is not smaller than a constant, it can be shown that a corner equilibrium in a structure is Pareto efficient for this given structure. As there is only one final good in our model, this implies that the corner equilibrium in structure D maximizes the per capita output level of the final good in this structure. This, combined with a fixed population size, implies that the maximum output level of the final good in structure D is given by the corner equilibrium in structure D.

Assume that there is no transaction cost, i.e. $k = t = 0$ in (3b) and (7); a comparison of the maximum per capita output level of final good (which equals the per capita real income) between structures D and A yields the result that the system of production functions (1) exhibits economies of division of labor if

$$F = a(1 - a)^{1/a - 1}(1 + b)^{1+b} / b^b > 1,$$

(13)

where $\partial F / \partial a > 0$ and $\partial F / \partial b > 0$. This leads us to

Proposition 3. The system of production functions (1) exhibits economies of division of labor if either (i) economies of specialization in producing the final good are sufficiently great, provided that $b$ is not too small, or (ii) economies of specialization in producing the intermediate good are sufficiently great, provided that $a$ is not too small.

The implications of expression (13) are that economies of specialization in producing a single good (either the final or intermediate good) are neither necessary nor sufficient for the existence of economies of division of labor. Economies of specialization in producing a single good are not sufficient for the existence of economies of division of labor since we can find some $a > 1/2$ (i.e. economies of specialization exist in producing the final good) and a sufficiently small $b$ such that (13) does not hold (i.e. no economies of division of labor). However economies of specialization in producing both goods are sufficient for the existence of economies of division of labor because (13) must hold if $a > 1/2$ and $b > 1$. 
Economies of specialization in producing the final good are also not necessary for the existence of economies of division of labor because we can find some \( a < 1/2 \) (i.e. economies of specialization do not exist in producing the final good) and a sufficiently large \( b \) such that (13) holds (i.e. economies of division of labor exist). In addition we can find some \( b < 1 \) (economies of specialization do not exist in producing the intermediate good) and a value of \( a \) that is sufficiently close to one (i.e. economies of specialization in producing the final good exist) such that \( a < 1 \) and (13) holds. This implies that the input–output relationship between goods provides the possibility that a sector with a significant degree of economies of specialization can promote productivity in other sectors which have no economies of specialization. In structure D, if \( b < 1 \), i.e. economies of specialization do not exist for \((x/y)\) (or \( a < 1/2 \), i.e. economies of specialization do not exist for \((y/x)\)) the economies of division of labor are external to configuration \((x/y)\) (or \((y/x)\)). However, if economies of specialization do not exist in any configurations, economies of division of labor cannot arise because (13) cannot hold if \( a < 1/2 \) and \( b < 1 \). In other words, economies of division of labor cannot be external to all individuals in our model. This provides a formal underpinning for Knight’s argument (1925) that the concept of increasing returns that are external to all firms is an ‘empty economic box’. On the other hand, even though economies of division of labor may be external to some individuals, our model will show that a decentralized market can trade off economies of specialization against ex ante or anticipated ex post transaction costs in order to exploit such externalities and diffuse the economies of specialization which exist in one configuration to other configurations which do not have economies of specialization. This is an important function of a free market system.

In order to see this function, we have to show that a general equilibrium is Pareto optimal. Since a corner equilibrium is a restricted Pareto optimum for a given structure, and a general equilibrium is the corner equilibrium with the maximum per capita real income, a general equilibrium is certainly Pareto optimal. This, together with corollary 1, implies that a decentralized market will fully exploit economies of division of labor if these economies outweigh the transaction costs.

If there are economies of division of labor and there are no transaction costs, the division of labor can be organized through markets for final and intermediate goods. The institution of the firm is not required. However, if there are transaction costs and there exist economies of division of labor, the free market will search for the most efficient organizational structure in transactions among structures D, FY, and FX. In order to highlight the conditions for the emergence of the institution of the firm, we need to define economies of the firm. The system of production (1) and transaction technology is said to display economies of the firm if the per capita real income is greater in structure FY or FX than in structures A and D. This definition, combined with propositions 1 and 3, and the comparison of per capita real income across the various structures yields.
Corollary 2. No economies of the firm exist if there are no transaction costs. Economies of division of labor between production of the final good and production of the intermediate good are necessary but not sufficient for the existence of economies of the firm. There exist economies of the firm if economies of division of labor outweigh the transaction cost in structure FY (or FX) and the transaction efficiency is lower in trading the intermediate good than in trading the labor hired to produce the intermediate good.

Coase (1937) surmises that economies of division of labor are not sufficient for the existence of economies of the firm. It is also necessary, he argues, that transaction costs are positive. Corollary 2 has formalized and refined these ideas. Cheung (1983) has developed Coase's theory of the firm by pointing out that the firm is replacing a goods market with a labor market rather than replacing market with non-market institutions. Our model has formalized the idea of Cheung. In addition, our model can be used to verify the Coase theorem (1960) if we interpret the Coase theorem in a special way that relates to our model.

In the model presented in this paper, the structures of production and the division of labor are the same, but the structures of residual rights differ between market structures D, FX, and FY. In structure D, no employer-employee relationship exists, or in Grossman and Hart’s words (1986), authority and residual rights of control and returns are equally distributed between all parties to a contract. In market structure FX or FY, authority is asymmetrically distributed between an employer and his employees. The employer claims the residual rights to returns and control of labor. If we identify the ownership of a firm with the claim to residual rights of a firm, then there are two structures of ownership of a firm in our model. In structure FX a producer of x claims residuals and owns the firm. Alternatively in structure FY a producer of y claims residuals and owns the firm. Corollary 2 implies that an asymmetric distribution of residual rights to returns and control is more efficient than a symmetric distribution of residual rights to returns and control if economies of the firm exist. Proposition 1 and corollary 1 state that the equilibrium must be one of structures FX and FY if economies of the firm exist. Together with our discussion of the Pareto optimality of any equilibrium, this implies that the structure of ownership of a firm will affect the equilibrium and the Pareto optimum. This conclusion is the same as in Grossman and Hart (1986) which shows that the structure of ownership matters. However, our focus differs from that of Grossman and Hart (1986) and Hart and Moore (1990) who emphasize the role of specificity of tangible assets in understanding the implications of a structure of ownership for the theory of the firm. Consistent with the Coase theorem, our model shows that the structure of ownership makes no difference if transaction cost is absent, but the structure of ownership may matter if transaction costs are not zero. However, our model shows that the existence of transaction costs is necessary but not sufficient for the equilibrium implications of the structure of ownership. For instance, structures D, FX, and FY
generate the same per capita real income, so that the structure of ownership makes no difference if \( t = u \) and \( h(t, r, k, a) = 1 \), where \( h(t, r, k, a) \) is given in footnote 11.

The implication of the analysis in this paper is that claims to residual rights can avoid the pricing of efforts of the claimant of residual rights, thereby reducing transaction costs if such pricing is prohibitively expensive. Let us use an example to illustrate this point. Suppose \( y \) is a tangible consumer good, such as cloth and \( x \) is an intangible management service in producing cloth. It is prohibitively expensive to measure and price \( x \) or the labor used to produce \( x \), or to measure the relationship between efforts exerted producing \( x \) and the quantity and quality of \( x \). For instance, it is impossible to tell if a manager is thinking about management strategy or about his girl friend when he is sitting in office. This assumption is equivalent to a large value of \( t \), the transaction cost coefficient of the intermediate good, and a large value of \( u \), the transaction cost coefficient of labor hired to produce \( x \). Assume, on the other hand, that it is very cheap to price labor hired to produce \( y \), because it is easy to measure the relationship between efforts exerted producing \( y \) and the quantity and quality of the tangible final goods. For instance, the quality and quantity of cloth produced by a worker is easy to measure and it is easy to tell that the worker is shirking if he does not move his hands. This assumption is equivalent to small values of \( k \), the transaction cost coefficient for trading the final good, and \( r \), the transaction cost coefficient for trading labor hired to produce \( y \). For the structure of residual rights in market structure FX, the final good \( y \) is exchanged for labor hired to produce \( y \). Hence, good \( y \) and efforts exerted producing \( y \) have to be priced, but the producer good \( x \) and efforts exerted producing \( x \) need not be priced in this structure. Therefore, this structure of residual rights is a special way of replacing the pricing of intangible \( x \) and the efforts exerted in producing \( x \) with the pricing of the tangible good \( y \) and the effort exerted in producing \( y \).

The claim to residuals of a firm owned by the producer of \( x \) is an indirect price of the employer's effort. This indirect pricing of management efforts via claims to residual rights avoids the prohibitively high costs in pricing \( x \) and in pricing the labor used to produce \( x \). Under our assumptions of the particular value ranges of parameters, proposition 1 and corollary 1 imply that structure FX is the equilibrium and is Pareto efficient. That is, under our assumption of the value range of transaction cost coefficients, a producer of management service owns a firm and claims its residuals in the equilibrium where there are markets for the consumer good and for the labor used to produce consumer good \( y \) and there is no market for the management service and for the labor used to produce the management service \( x \). Therefore, this market structure avoids the problem of pricing management services and the labor used to produce management services. Under our assumptions, structure FY is not efficient and cannot be an equilibrium because it involves a market for the labor used to produce the management service which is associated with a prohibitively high transaction cost. Structure D is not efficient.
and cannot be an equilibrium because structure D involves a market for management services which is associated with a prohibitively high transaction cost.

Intuitively, the division of labor in producing producer and final goods not only generates transaction costs, but also creates more scope for economic players to choose a method of organizing transactions in order to save on transaction costs. The method of organizing transactions can be a two combination of four factors. The four factors are: trading the final good, trading the producer good, trading the labor used to produce the final good, and trading the labor used to produce the producer good. The first feasible combination of the four factors is trading the final and producer goods (structure D); the second is trading the final good and the labor used to produce the producer good (structure FY); the final is trading the final good and the labor used to produce the final good (structure FX). An equilibrium structure of residual rights will avoid pricing output (good) and input (labor) of the activity that has the lowest transaction efficiency.

The implications of this example for productivity progress and for the evolution in the division of labor are straightforward. Suppose that the transaction cost coefficient of the management service, \( t \) and the transaction cost coefficient of labor used to produce management service, \( v \), are extremely large, then when structure FX is not allowed the division of labor will generate a lower per capita real income (utility) than autarky. The emergence of structure FX will allow the division of labor to generate a higher per capita real income than in autarky. This implies that a particular structure of residual rights may promote the division of labor and productivity via improving transaction efficiency. The structure of residual rights achieves this by allowing individuals to specialize in the production of \( x \) while avoiding the direct pricing and marketing of \( x \) and related efforts. This analysis yields the following corollary:

**Corollary 3.** The structure of residual rights and the structure of ownership of a firm are critical for the determination of equilibrium and the Pareto optimum if transaction cost differs across goods and factors. Provided economies of the firm exist, structure FX is the equilibrium if transaction efficiency of labor hired to produce the final good is higher than that hired to produce the intermediate good; otherwise structure FY is the equilibrium.

Although a structure of residual rights associated with the firm can avoid the direct pricing and marketing of the activity that has the lowest pricing efficiency or transaction efficiency, only one activity can be excluded from direct pricing and marketing. If there are many goods and services and increasingly more goods are traded as transaction efficiency is improved, then the activity which is indirectly priced via claims to residual rights of a firm may be changed. For instance, the producer of management service is the claimant of residual rights before the division of labor between production management and portfolio management emerges. The portfolio managers, or shareholders may become the claimants of
residual rights as the division of labor between production management and asset portfolio management emerges. This can be explained by a higher transaction cost of pricing efforts of portfolio management than transaction costs of pricing efforts of production management and by the fact that only one activity with the lowest transaction efficiency can be excluded from direct pricing and marketing by a structure of residual rights.

This story cannot be generated by simply assuming different values of the transaction cost coefficient in structure D since even if the transaction cost coefficient is the same for structures D, and FX, structure D will generate a higher percapita real income due to a differential in the structure of transactions between D, where the specialist producers of the final good self-provide the final good, and FX, where the specialist producers of the final good have to buy the final good from the market. The theory of the firm developed in this paper will be referred to as 'theory of indirect pricing' and the theory of the firm developed by Grossman, Hart, and Moore will be referred to as 'theory of asset specificity'. A complete story of the firm that occurs in reality may be then predicted by a blend of the theory of indirect pricing and the theory of asset specificity. It is a useful exercise to extend this model to the case with several final goods; this is undertaken in Yang (1988b).

5. Implications and conclusions

In this paper, a framework with consumer-producers, economies of specialization, and transaction costs has been used to explore the role of a structure of residual rights that is associated with the institution of the firm. An asymmetric structure of residual rights can be used to improve transaction efficiency and to promote the division of labor by excluding the activity with the lowest transaction efficiency from direct pricing and trading. In other words, a structure of residual rights that is associated with a firm may get an activity whose outputs and effort inputs are intangible involved in the division of labor but avoid direct pricing and marketing of this activity. In this way, the institution of the firm plays a role similar to that of a patent law through improving pricing efficiency of intellectual properties. According to the theory of indirect pricing, legislation of a law that protects free association and the residual rights of owners of firms is a driving force in economic development. The intuition behind the theory is straightforward. If an entrepreneur has a good idea to develop a new business to make money but it is extremely expensive to sell this idea in the market because it is extremely expensive to enforce his rights to the intangible intellectual property, then the best way for him to develop the business meanwhile enforcing his intellectual property rights is to set up a firm and to hire workers to do whatever he wants them to do. By doing so, he indirectly sells his intangible intellectual property to the market while avoiding its direct pricing and marketing.
The theory developed in this paper is motivated by two observations. First, it can be used to resolve a puzzle in Chinese economic history. Elvin (1973) has documented that Song China (960–1270 A.D.) possessed both the scientific knowledge and the mechanical ability to have experienced a full-fledged industrial revolution some four centuries before it occurred in Europe. Chinese developed very elaborate contracts and sophisticated commercial organizations at that time. However, all Emperors were extremely sensitive to unofficial free associations because dissidents tended to use such associations to develop underground anti-government movements due to the characteristics of the dynasty cycle in Chinese history. Hence, there was no legal system which protected the residual rights of entrepreneurs to any manufacturing firms. Instead, the government tended to infringe arbitrarily upon such residual rights. This discouraged investments especially in manufacturing industries. According to the theory of the firm developed in this paper, it is the absence of a legal system which protects residual rights to firms, compounded by the absence of patent laws, that prevented China from expanding its technical inventions to large scale commercialized production, which might have brought about an industrial revolution.

The second observation is related to the development implications of the absence of a legal system that protects residual rights to firms in a Soviet style socialist economy. Due to Marx’s ideology, such residual rights are considered as the source of exploitation in a Soviet style socialist country. Also, a Soviet style government prohibits unofficial free association (including free enterprises) for political reasons. According to our theory of indirect pricing, it is this institutional arrangement rather than distorted relative prices of tangible goods that is responsible for the shortage of innovative entrepreneurial activities in such an economy.

Appendix. Proof of proposition 1

There are $2^6 = 64$ combinations of zero and non-zero values of $x$, $x^s$, $x^d$, $y$, $y^s$, $y^d$. Applying the Kuhn–Tucker theorem, we may rule out many of the combinations from the list of candidates for an individual’s optimum decision. We need not count these combinations in solving for equilibrium. Using (1), (2), and the budget constraint

$$x^s + py^s - x^d + py^d;$$

an individual’s utility can be written as

$$U = \left\{ \left[ (1 - L_y)^b - x^s + (1 - t) x^d \right] L_y \right\}^a + (1 - k) y^d - \left( x^d - x^s + py^d \right)/p, \quad \text{if } y^s > 0,$$

$$U = \left\{ \left[ (1 - L_y)^b - x^d - py^d + (1 - t) x^d \right] L_y \right\}^a + (1 - k) y^d, \quad \text{if } y^s = 0.$$

\[\text{(A-2a, A-2b)}\]
Differentiating (A-2), it follows
\[ \frac{\partial U}{\partial y^d} < 0 \quad \text{if} \quad y^s > 0 \] (A-3a)
\[ \frac{\partial U}{\partial x^d} < 0 \quad \text{if} \quad \frac{\partial U}{\partial x^s} = 0 \quad \text{and} \quad y^s > 0 \quad \text{or} \quad x^s > 0 \quad \text{and} \quad y^s = 0 \] (A-3b)

According to the Kuhn–Tucker condition \( \frac{\partial U}{\partial y^d} y^d = 0 \) at the optimum, (A-3a) implies that the optimum \( y^d \) is zero if \( y^s > 0 \). Similarly, (A-3b) implies that the optimum \( x^d \) is zero if the optimum \( x^s > 0 \). Assume \( x^s \) and \( y^s \) are positive at the same time, then (A-3) implies that \( x^d = y^d = 0 \) which contradicts the budget constraint because a person will not sell if he does not buy. In summary, (A-3) and the budget constraint implies that an individual sells only one good and does not buy and sell a good at the same time and thereby the optimum decision is

\[ \begin{align*}
\text{either} & \quad x^d = y^s = 0, \quad x^s, \quad y^d > 0, \\
\text{or} & \quad x^d, \quad y^s > 0, \quad x^s = y^d = 0, \\
\text{or} & \quad x^d = x^s = y^s = y^d = 0,
\end{align*} \] (A-4)

where (A-4c) is autarky; (A-4b) is configuration \( (y/x) \) except we are not sure \( y > 0 \); (A-4a) is configuration \( (x/y) \) except we are not sure \( x = L_y = 0 \). Therefore we can prove that the optimum decision is one among configurations autarky, \( (x/y) \), and \( (y/x) \) if we have shown that \( y > 0 \) if (A-4b) is true and that \( x = L_y = 0 \) if (A-4a) is true.

Assume (A-4b) holds true; the quantity consumed of the final good \( y + (1 - k)y^d \) is zero if \( y = 0 \). Therefore a positive utility level implies

\[ y \text{ must be positive if (A-4b) holds true.} \] (A-5)

In other words, we rule out combination of \( x^d, \quad y^s > 0 \) and \( y = 0 \) from the list of candidates for the optimum.

Assume that (A-4a) holds true and note (1), (2) and (A-1); (A-2) can be written as

\[ U = [xL_y]^a + (1 - k)[(1 - L_y)^b - x]/p. \]

Assume \( x > 0 \); the optimum value of \( x \) is given by

\[ \frac{\partial U(x, L_y)}{\partial x} = 0. \] (A-6a)

Inserting this optimum value of \( x \) into (A-5) and differentiating \( U \) with respect to \( L_y \), we have

\[ dk\{ \frac{\partial U[L_y, x(L_y)]}{\partial L_y} / dL_y \} > 0 \quad \text{if} \quad \frac{\partial U(x, L_y)}{\partial L_y} = 0, \]

\[ 0.5 < a < 1 \quad \text{and} \quad b > 1 \] (A-6b)

where \( x(L_y) \) is given by (A-6a). It can be shown that \( dk\{ \frac{\partial U[L_y, x(L_y)]}{\partial L_y} / dL_y \} > 0 \) if the Hessian of \( U(x, L_y) \) with respect to \( x \) and \( L_y \) is negative. This, combined with (A-6b), implies that the interior extreme of \( L_y \) is not the maximum
and the optimum $L_y$ may be at a corner, i.e. either $L_y = 0$ or $L_y = 1$ if $x > 0$, $0.5 < a < 1$ and $b > 1$. $L_y = 1$, i.e. $L_x = 0$ contradicts the assumption of $x^s > 0$. Thus,

the optimum $L_y = 0$ if $x > 0$. \hfill (A-6c)

Since

$$\frac{\partial U}{\partial x} < 0 \quad \text{for any} \quad x \quad \text{if} \quad L_y = 0,$$ \hfill (A-6d)

the optimum $x$ is zero if $L_y = 0$. This contradicts (A-6c). Hence, a positive optimum value of $x$ is impossible if (A-4a) is true. Assume that $x = 0$; it is easy to see that the optimum $L_y = 0$ because $\frac{\partial U}{\partial L_y} < 0$ for any $L_y$. Hence, (A-6) means

$$L_y = x = 0 \quad \text{if} \quad (A-4a) \quad \text{is true,} \quad b > 1, \quad \text{and} \quad 0.5 < a < 1. \quad \text{(A-7)}$$

(A-4,5,7) lead us to proposition 1 which implies that the list of candidates for the optimum consists of configurations autarky ($x/y$), and ($y/x$). Similarly, we can prove that the list consists of the seven configurations in Section 3 if we take into account the possibility for establishing firms.

References


Williamson, Oliver, 1975, Markets and Hierarchies (Free Press, New York).
Williamson, Oliver, 1985, Economic Institutions of Capitalism (Free Press, New York).