Why Firms and Markets in Economics?

Sung-Hee Jwa

In this paper, by carefully reflecting on the modern transaction costs approach to the Theory of the Firm, we try and understand better the economic nature of firms vis-a-vis markets to derive important corporate policy implications. By following the idea of Coase that firms come to existence to economize on transaction costs, we specifically put it that the firm does so by internalizing factor market rather than product market activities. We also establish that the firm is essentially characterized by its vertical structure where the coordination function is realized through 'command and control,' while markets are horizontal, democratic and spontaneous. This paper goes further to redefine the firm as a collection of non-transparent economics activities. Although the term 'transparency' has been commonly used in the media and in policy debates, its precise economic interpretation has been so far unsatisfactory in the economics literature. This paper attempts to provide an economic meaning of transparency, which is placed in the context of firms and markets. Various corporate issues including its governance, policy, and transparency, as well as the role of market disciplinary mechanisms are clarified through theoretical extensions and discussions regarding the theory of the firm.

Keywords: Theory of the firm, Nature of firms and markets, Transaction costs, Factor and product markets, Corporate governance, Corporate transparency

JEL Classification: D21, D23, D40

*President, Korea Economic Research Institute, FKI Building, 12 Fl. 28-1, Yoido-dong, Yeongdeungpo-ku, Seoul 150-756, Korea, (Tel): +82-2-3771-0004, (Fax) +82-2-765-0270/1. The views and opinions of the author expressed herein do not necessarily state or reflect those of the Korea Economic Research Institute.

I. Introduction

This paper critically assesses the recent theoretical advancements in the theory of the firm and specifically ‘transaction costs’ economics, in order to provide a better understanding of the nature and roles of firms and markets in economics. Albeit the simplicity of the research question, a closer look reveals that economists are still far from providing a non-controversial theory of firms and markets. This paper has a dual purpose of tying up some loose ends, while adding to controversy.¹

Looked at one way, traditional microeconomics contains many technical refinements concerning the market and firms. Looked at another way, most formal models of the firm are very rudimentary and highly abstract bearing little relation to the real world complex firms, and as such, in terms of practicality, very little has been learnt. The firm in neo-classical economics was for a long time described by a technical ‘black-box’ explaining production activity using production and cost functions—most economists relegated the study of the internal workings of the firms to business schools as it was thought not worthy of Economic Science. The views of economists have changed increasingly over the past two decades, as many scholars find themselves returning to Coase’s (1937) seminal paper ‘The Nature of the Firm.’

After Coase (1937), there is now a variety of different schools, some related some not, each making important theoretical advancements in our understanding of firms and markets. Particularly worth mentioning, in the neoclassical lines on contract theory is the Agency-Principal approach of Alchian and Demsetz (1972), Stiglitz (1974) and Jensen and Meckling (1976); The ‘Transaction Costs Economics’ approach associated with Williamson (1975, 1985) is well-known; There is also the pioneering work of Klein, Crawford, and Alchian (1978), and the modern property rights approach of Grossman and Hart (1986) and Hart and Moore (1990); the resource-based approach of Penrose (1958) is also important as well as the more recent thinking stimulated by the evolutionary

¹Ideas in this paper were included in my ‘Opening Remarks’ for the 2002 Seoul Journal of Economics International Symposium of the title ‘The Socio-Economic Roles of Firms and Chaebols.’ I would like to acknowledge the excellent research assistance of Yong Yoon.
approach of Nelson and Winter (1982). Although I have listed here
some of the more important works that any serious economists
should be familiar with, in this paper, I do not attempt to survey
all existing literature but rather wish to focus on the difference
between the nature of firms and markets in economies, and derive
some important policy implications. Specifically, I wish to see how
recent development in the theory of the firm helps us to
understand the boundaries of the firm, and by doing so, I shall not
only try to solve various contradictions, but will also discuss
relevance and implications to real world issues in corporate
governance, public policy and development economics.

II. Separating Firms and Markets

Any serious discussion on the firm and markets must begin with
Coase (1937), which poses two basic questions: Why firms exist?
And, what determines the scale and scope of firms (the boundary
of the firm)? Coase’s answer was provided in terms of transaction
costs incurred in transacting in the world of imperfect information.
He writes, ‘a firm will tend to expand until the costs of organizing
an extra transaction within the firm become equal to the costs of
organizing an extra transaction by means of an exchange on the
open market or the cost of organizing in another firm.’ Put simply,
while retaining the original meaning, when transaction costs of
market exchange are high, it may be less costly to coordinate
production through a formal organization or firm than through
markets.

In a similar vein, in an excellent article on the theory of the
firm, Alchian and Woodward (1987) puts it succinctly that firms
come to existence because of transaction costs – costs associated
with determining quality and negotiating prices. Transaction costs
are associated with defining and enforcing terms of trade. Imperfect
information, as well as human limitations such as bounded
rationality, makes it difficult to properly define terms of trade or
fully enforce them. This non-transparency of terms of trade lead to

\(^{3}\)Of course, there are other contributions in the area of game theory,
experimental economics and so on. An excellent survey of the theory of the
firm is Holmstrom and Tirole (1989).

\(^{4}\)Coase (1937, p. 395).
opportunistic behavior such as hold-up problems, agency problems, moral hazard behavior and so on, all of which add to higher transaction costs. The firm we are told therefore comes into existence minimizing or even avoiding such transaction costs. Given that activities that have been knocked-out of the market and integrated into firms are precisely those of higher transaction costs, non-transparent economic activities, I state the first premise of this paper, which is also a redefinition of the firm, as follows:

**Premise 1:** A Firm Is a Collection of Non-Transparent Economic Activities.

That is, we claim that if it is a firm, then it must be non-transparent. Under the logical statement, ‘if $P$, then $Q$’ or $(P \Rightarrow Q)$, we can construct its contrapositive, that is ‘if not $Q$, then not $P$,’ that is $(\neg Q \Rightarrow \neg P)$, which is always true if the original statement is true. In this case, we can argue, if not non-transparent (that is, transparent), then not a firm. In the context of our discussion, by ‘not a firm,’ we are actually referring to perfectly competitive markets. It is not incorrect to claim that a perfectly competitive market then is a collection of transparent economic activities. However, we should bear in mind that perfectly competitive markets do not exist in reality. In essence, we can only talk of a continuum of degrees of transparency with firms and markets on opposite ends of a transparency, or if you like, non-transparency line.

Another crucial question regarding firms and markets may be asked: how are firms able to do minimize or even avoid transaction costs? What is that element or characteristic that firms have, which markets do not, that allows firms to enjoy lower costs for certain high transaction cost activities, and on top of that, still maintain economic coordination? The answer, I suggest, is ‘flat’ – or if you like ‘coordination by fiat’ – which explains the hierarchical coordinating structure of firms that contrasts sharply with the horizontal structure of market transactions. Accordingly, a second premise

---

1At this point, the reader familiar with the theory of incomplete contracts may appreciate this point better. Nonetheless, what I wish to emphasize is the nature of transparency rather than the fact that contracts may never be complete.
that distinguishes the firm from markets is as follows:

**Premise 2:** A Firm Replaces the Horizontal Structure of Market Transactions with Command and Control.

The importance of understanding the differences between firms and markets cannot be overemphasized. What needs to be clearly understood is the tension, as Rosen (1991) puts it, ‘between the use of the price system to elicit spontaneous coordination and the recourse to hierarchy to effect conscious coordination.’ We have shown that a firm operates within a market but is also a negation of the market, and that the firm and markets are different with regards to their degree of transparency and coordination structure.

III. **Contradiction Between the Northian and Coasian Interpretations of Transaction Costs**

It is interesting to note that the approach by Coase, Williamson, and other economists of the transaction costs line have largely missed out in their arguments of the relationship between transaction costs and economic development. While having presented penetrating analysis and insights as to the implication of transaction costs at the margin of firms and markets, little has been said of transaction costs in a macro sense – that is in terms of the overall economy rather than to a specific single product, firm or market. An exception can be found in the work of Douglass North (1992). We shall now see how different interpretations of transaction costs in economics can lead to theoretical contradictions, and suggest how they may be resolved.

Specifically, what is curious is that, upon closer inspection, Coase’s thinking suggests that a firm will expand on the margin as

\[\text{Interestingly, among neo-classical economists, the closest to relating transaction costs to economic performance can perhaps be credited to Arrow (1969, p. 48) who identified market failures as having transaction costs origins and ‘which in general impede and in particular cases block the formation of markets,’ – Arrow defines the concept of transaction costs broadly as ‘costs of running the economic system’ (1969, p. 48). However, neo-classical economics has not been able, for analytical limitations presumably, to directly link transaction costs or any other mechanism to the process of economic development.}\]
it 'internalizes' higher transaction cost activities that the firm is better off absorbing. This I have occasionally termed the Coasian effect—the size of the firm will tend to increase as it battles with higher transaction costs.

North (1990, p. 67) writes that, 'Firms come into existence to take advantage of profitable opportunities, which will be defined by the existing set of constraints. With insecure property rights, poorly enforced laws, barriers to entry, and monopolistic restrictions, profit-maximizing firms will tend to have short time horizons and little fixed capital, and will tend to be small scale. The most profitable businesses may be in trade, redistributive activities, or the black market. Large firms with substantial fixed capital will exist only under the umbrella of government protection with subsidies, tariff protection, and payoff to the polity — a mixture hardly conducive to productive efficiency.' Thus, North has argued in relation to property rights protection, that we can expect larger firms only when property rights are secure — that is, under low (not high) transaction costs. I have termed this the Northian effect.

The puzzle then is that the Coasian and Northian approaches seem to carry contradictory theoretical implications, particularly as relates to the size of the firm. With larger (lower) transaction costs, the former argues that firm size tends to be larger (smaller), while the latter suggests firm size to be smaller (greater).  

IV. Factor Markets Versus Product Markets: Which Are Internalizable?

One way out of this contradiction is to differentiate between factor markets and product markets.  

6The debate of what determines firm size, which can be traced back to the 1930s in the work of a French economist Robert Gilbert, is neither trivial nor uncontroversial. An interesting recent paper is Zingales et al. (1999).

7This section draws from Chapter 7 in Jwa (2002). In this paper, we may assume that markets are distinguishable by placing them under the general context of Williamson (1981, p. 1544), 'technologically separable goods and services,' (i.e. a phase where one stage of processing or assembly activity terminates and another begins). This technical concept has been useful for Williamson in defining transaction across distinguishable markets, which we borrow to allow for the separation of markets, specifically, factor and
refers to those markets where resources for the production of goods and services are made available, while product market means markets where a firm’s final goods and services (production) are sold. Be warned that the same market can be a factor market to one firm while at the same time be a product market to another – care must be observed when making this distinction between factor and product markets. For example, consider the financial sector: A securities company may sell financial services in the product market, say, to some manufacturing company, but this same product market, in the perspective of the manufacturing firm is actually a factor market where financial resources can be purchased. The Coasian effect essentially concerns factor markets – a firm is an organization that ‘internalizes’ factor markets – hence, larger firms result from efforts to tackle higher transaction cost activities in factor markets. On the other hand, the Northian effect concerns the overall product market or end-market if you like. To make this argument formally, we introduce the third premise as follows:

**Premise 3:** A Firm Is a Negation of Factor Markets – It Internalizes or Absorbs Higher Transaction Cost Activities in Factor Markets.

What is important to realize is that transaction costs in the product market by definition cannot be absorbed into the firm, but rather act as an absolute constraint to the firm – selling products to yourself doesn’t constitute a market in the ordinary sense. Hence, we find that transaction costs in factor markets and product markets differ in their implications on firm size – the former type of transaction costs are internalizable and in the Coasian sense tend to add to firm size, while the latter transaction costs cannot be absorbed and thereby acts, in the Northian sense, as a constraint limiting the growth of firms.\(^8\)

The distinction I make between the factor and product markets and their relationship to the firm is not only useful in resolving the theoretical contradiction to firm size of the Coasian and Northian product markets.

\(^8\)Coase (2002), while speculating on the impact of the Internet on firm size, discusses briefly yet intricately about the subtle effects of transaction costs on firm size not dissimilar to the arguments formalized in this paper.
approach, but also has further implication in helping us understand why traditional corporate governance may not be as effective as originally expected. This is discussed in Section VI.

V. Transparency in Corporations and Markets

Corporations have come under attack more often than is due—true, there are many things that certain firms have done that do not seem ‘right,’ but many of the criticisms are often unwarranted simply because of our improper understanding of the economic roles of corporations and markets. How should we interpret the realities of economic issues concerning corporations given the above arguments regarding the theory of the firm? To make things more concrete, let’s take the case of corporate America. Recently, the opinion of nearly every major finance and economic newspaper in the past months—after Enron, Global Crossing, WorldCom and Tyco—is that corporate America is no longer viable or to be trusted. This is hardly anywhere near the truth. Although such big corporate names have recently been exposed, it does not necessarily follow that corporations on the whole are set to failure. I provide here my explanation for the current crisis in corporate America.

First, I think that the ‘pendulum has swung too far’ in favor of markets, and firms in recent times have been losing far too much ground. The firm, as I have defined, is a collection of non-transparent economic activities (see Premise 1) and the investor, in fact, incurs a risk but is rewarded by profit precisely because of the non-transparent nature of firms. Since the difference in transparency is one of the distinguishing features of firms and markets, demanding increased transparency on the part of firms to match that of markets is unwarranted. For example, regulations requiring forced disclosure of propriety information can discourage innovation, as firms occasionally rely on being non-transparent to ensure innovators are rewarded for their efforts.9 If a firm is forced into transparency to the extent of markets then it inevitably

9In a similar vein, Alan Greenspan (2002), questioning the desirability of unconditional transparency, put it that although ‘there should not be much dispute that markets function best when the participants are fully informed ... paradoxically, the full disclosure of what some participants know can undermine incentives to take risk, a precondition to economic growth.’
narrow the potential contribution of firms to the economy by eroding the very reason for their existence. This issue of firm transparency is hardly trivial and there are many technical difficulties that make it difficult, and as we have argued, improper to try and blatantly make firms transparent. For example, the GAAP (Generally Accepted Accounting Principles) has recently been proposed as THE common set of accounting principles, standards and procedures that ALL firms are expected to follow. But, I'm afraid that there cannot be a single system that makes transparent all terms of trade of internal transactions in corporations. For instance, think of asset or investment valuation – these, like many accounting practices, unavoidably rely on a high degree of subjectivism with much of it based on value judgment and all decisions cannot be fully objective. So, the GAAP can be only one of many possible accounting systems, each having its own strengths and weaknesses.

Furthermore, in recent years, even in countries like the US where investor protection is already well established, there has been increased pressure for transparency especially from large investors as their market power is strengthened enabling them to push investment risk to be borne by firms.\textsuperscript{10} This can only hurt investors and firms, as well as the whole economy. While transparency is desirable in general, it is unreasonable to demand that the firm's transparency match that of markets, as this destroys the essence of firms – it is in this sense that I say that the pendulum seems to have swung too far and balance needs to be restored.

Although my argument so far may seem to favor corporations, it should be noted that in no way do I diminish the importance of corporate transparency and of certain constraints (both legal and informal or societal) on management decisions. Although, as mentioned earlier, the firm is a collection of nontransparent economic activities, there remains the question of what levels of non-transparency may or may not be optimal to firm and economic growth. We should take note from the outset that the degree of corporate transparency that is optimal is not fixed for all corporations, but on the average are subject to the degree of market imperfections, which are, of course, a reflection of societal

\textsuperscript{10}For a good survey of the trend of impact of institutional investors in the US see Holmstrom and Kaplan (2001).
overall institutional factors, both formal and informal, and which differ from country to country and across individual economies. In short, the level of corporate transparency vis-à-vis markets, and the level of investor protection as well evolve endogenously according to societal needs. However, certain acceptable ‘rules of the game’ should exist as part of the institutional setting that allow for both firms and investors to interact to bring about social benefits, for example, in order to attract sufficient capital from outside investors, as well as properly reward their investments. That is, as an important part of a nexus of institutional factors influencing firm behavior, certain requirements may exist demanding a minimum degree of corporate transparency, which if absent can seriously discourage outside investment and result in sub-optimal growth of firms. Nonetheless, it remains that one should always be suspicious of demanding ‘excess’ transparency that is antagonistic not only to innovativeness, but also to the firms’ *raison d’être*. Crisscrossing and blurring the nature of firms and markets dissolves the essence and vital economic functions of firms and markets. We summarize our point by the fourth premise as follows:

**Premise 4:** Transparency Is to Markets as Non-Transparency Is to Firms.

**VI. The Failure of the Traditional Corporate Governance Systems**

One relevant question is then to explain why Corporate America has been driven into the current chaos despite its corporate governance system, hailed as THE best in the world? Why doesn’t the world’s best corporate governance system\(^\text{11}\) seem to be as effective as intended? In what follows, I will identify two fundamental flaws.

Traditional corporate governance systems rely perhaps excessively on factor market pressure to discipline firm behavior.\(^\text{12}\) It follows

\(^\text{11}\)For those new to issues on corporate governance, see Shleifer and Vishny (1997), Vives (2000) and, in the context of East Asia, Jwa (2000) – an earlier separate article of mine in this journal.

\(^\text{12}\)Recall here the distinction between factor markets and product markets made in section IV.
from Premise 3 above, that a firm, because it exists out of the efforts to 'internalize' factor markets, is able to avoid market disciplinary pressure simply by absorbing or finding alternative sources to replace existing factor markets. For instance, a firm may use internal finance as a substitute for external finance, say, through internally accumulated capital or funding across subsidiaries. The Board of Directors (BOD) system was originally created to represent shareholders' interests under the separation of ownership and management in a firm, and often is thought to exert effective disciplinary pressure on corporate behavior. In reality, however, the BOD system is simply nothing more than one way of exerting factor market type pressure on corporate management, and as such, the corporation may avoid disciplinary pressure from the BOD system, particularly, if the BOD system becomes a burden to management decision, say, by adding to increased transaction costs—organizational innovation or seeking alternative means that render BOD pressure on corporate management less effective becomes a natural choice of the firm. This is exactly the case that has been observed with the financial scandals of recent times in corporate America where the BOD system turns out to be ineffective in disciplining top management including the CEO. Evidently, the outside directors as well have not been checked. Here, one should be reminded that only those markets that cannot be integrated into the firm, that is the product market, could act as truly disciplinary mechanisms. Therefore, perhaps, there is need to restructure the traditional corporate governance system after a careful examination of the nature of firms and markets.

Secondly, with Premise 2, we established that the firm is based on command and control, and this vertical structure is what gives the firm an advantage over markets in reducing transaction costs thereby resulting in economic gain. This hierarchal structure implies that members of the BOD cannot run the company TOGETHER with the CEO. A ship with two captains is bound to go nowhere. In fact, we observe that the best corporations are precisely those with strong leadership in their CEOs, and a strong and intact hierarchal structure that allows the CEO to effectively exercise command and control. That is, despite variegated organizational forms, and the possibility of hybrid organizations, throughout the history of capitalism, we observe that firms with stricter vertical structures tend to outperform and outlive their
more horizontal counterparts. Be as it may, to be even more specific, it should be acknowledged that the BOD system is bound to fail because if it works as intended it can only go against the intrinsic nature of corporations thereby weakening survivability by complicating unnecessarily the firm's command and control structure.

Recent legal innovations illustrate various attempts to patch up the fundamentally flawed BOD system where outside directors have become redundant. For example, we have witnessed in the 1990s a flood of remedies such as 'class action suits' or 'stockholder's suits' against outside directors. Such attempts at rescuing this desperate system, however, is only circular and is bound to get us nowhere!

Then one might ask what the alternative corporate governance system will look like, which, in particular, may replace the fundamentally flawed BOD system. Factor market pressure, as we have seen, has limitations in effectively disciplining corporate governance. Rather, corporate discipline through product markets seems the best alternative, and it is in this sense that competition as a discovery process is critical for the overall health and effectiveness of firms and the economy. That is, competition promotion policy in the final product market should provide the most effective discipline to corporate behavior, and should therefore receive its due emphasis when addressing corporate governance issues.

VII. Implications for Public Policy

The discussions in this paper on the nature of the firm and markets, as well as issues on transparency and corporate governance, have important policy implications. In recent times there has been pressure of various degrees to structure firms' management along more 'democratic' lines. For example, there is increasing demands to include workers in the management decision-making process. Also, it is asked that the CEOs' control be limited in proportion to the shares held, and that all firms should

14 See Chapter 8, Jwa (2002) for further discussions of competition promotion policy, as well as the importance of free entry and exit, bankruptcy procedures and so on.
have an independent and strong outside director system – which controls the CEO – and so on. Such ‘democratic’ rules and regulations that ignore, and if imposed over, the ‘hierarchical’ structure of firms may eventually prove to be detrimental to firm and economy.

It is also interesting to look at the above logic from the other direction. That is, what if like calling for increased democratization within firms, we place artificial vertical structures over horizontal markets, say, by government fiat or decree, as was the case with Korea’s past development strategy. Confucianism, which has many teachings in administration, governance, leadership, punishment, and so on, and which has been the dominant philosophy in many East Asian countries for centuries, has resulted in the parental government. Free-markets have more or less been ignored throughout the modernization process and often I have warned that this is a serious mistake, which should be rectified promptly. As is evident from the Asian experience, placing strong vertical structures in the market tends to result in all sorts of problems, including bureaucratic inefficiencies, moral hazard behavior, misallocation of resources, and so on.

We have also seen that a certain degree of transparency is important to encourage investment from external sources, but one must be careful not to destroy the nature of the firm. Furthermore, withholding information can be important in securing the advantage needed to succeed in certain markets and to ensure that innovators be rewarded for their effort. Be as it may, demanding corporate transparency to match that of markets will only hurt firms and the economy.

---

15This is related to the question raised by Coase (1937, p. 340) when he asked, ‘Why is not all production carried on by one big firm?’

16For those that wish to know whether there is an alternative Eastern philosophy that may accommodate a market approach see Chapter 2 in Jwa (2001) where Taoism, which speaks of individual liberty, autonomy, and spontaneous order – a philosophy more akin to the idea of free markets – is introduced.

17Greenspan (2002) gives concrete examples of how secrecy may be important in realizing social benefit form a real estate developer, and how regulations on disclosure may hurt innovation particularly in over-the-counter derivatives markets.
The main policy lesson from the above discussion can be summarized as follows:

**Policy lesson:** Do Not Turn the Country into a Large Firm — like the former Soviet republic or North Korea — and at the same time, Do Not Turn Firms into Markets, by calling for ‘excess’ transparency and democracy in firms as in markets. It is **putting the economy into peril by contradicting the intrinsic nature of firms and markets.**

**VIII. Theory of the Firm in a Comparative Institutional Setting**

Although we have focused so far on the fundamental question why firms and markets in economics? Another equally important question begs an answer: Why Korean firms and markets? Why U.S firms and markets? Why German and Japanese firms and markets? That is, why do we have a variety of corporate forms as well as differentiated business strategies and behavior across different countries in the world? In fact, Williamson (1991) put it that ‘the choice between firm and market organization is neither given, nor largely determined by technology, but mainly reflects efforts to economize transaction costs,’ and transaction costs are, he goes further to say, ‘preeminently a comparative institutional undertaking.’ This implies that the types of corporations and their strategy and behavior, whether we refer to, for example, the keiretsu (in Japan) or the chaebol (in Korea) is nothing more than the generic firm that has risen and grown out of the country-specific institutional environment.  

In a similar vein, Caves (1989, p. 1227-8) writes that if ‘transaction-costs efficiencies of firms and markets were independent of laws, cultural traits, and other distinguishing traits of nationhood, then we should expect the allocation between firms and markets to differ only inessentially from country to country. Yet, one’s eye falls upon certain national institutions that seem quite distinctive.’ Hence, there is need to understand the idiosyncratic nature of firms that is specific to the country’s cultural, historical, and institutional background, and this.  

---

18 A fuller treatment of this theoretical perspective can be found in Chapter 7, Jwa (2002).
believe, will be a most important area for future research if more is to be learnt about many interesting questions related to the firms and markets, as well as their roles to the economic development process.

IX. Concluding Remarks

Despite the enormous achievements in the economists’ understanding of the firm by important works of transaction costs economists and others, there is still a need to rigorously rethink over and over again the theoretical aspects of the theory of the firm. Such a theory is still far from reaching a consensus, but our continued study nonetheless remains indispensable. Secondly, the wisdom derived from such work should not be detached from real world phenomena and should as much as possible incorporate into their analysis the most essential national characteristics as well as other institutional factors. To be kept in mind always is that the firm is institution-specific, because it is created and is constantly shaped not in isolation but in the context of the external institutional environment with which it interacts constantly. It is from this argument that I stress the need to address the nature of corporation and markets in a comparative institutional setting, which should provide the most important clues to understanding the process of economic development.

(Received 29 August 2002; Revised 5 March 2003)

References


