Transaction Costs and the Substitution at Multiple Margins

De-Xing Guan^{*}

June 2012

<u>Abstract</u>

In this paper we argue that the economics of transaction costs could be analyzed through studying the *substitution at multiple margins* and the resulting *substitution structure* of institutions. The transaction costs are related to the ease of substitution when economic agents make decisions at more than one single margin. We propose to use Morishima elasticity of substitution as a measure of the ease of substitution. An identity is used to explain the equality between the increase in total transaction costs and the total dissipation of rents in a society. Perfect competition, monopoly, and Welfare Theorems can be interpreted by this Identity. We also use Taiwan's minimum wage as a case to illustrate the use of substitution structure. Because it is easier to measure rents than transaction costs, and with the help of the Identity, we can use rents as a proxy variable for transaction costs, and this makes empirical works on transaction costs more promising. A more formal model is needed to address the empirical studies in this direction.

^{*} Department of Economics, National Taipei University, 151, University Road, San-Xia, New Taipei 23741, Taiwan. Email: <u>guan@mail.ntpu.edu.tw</u>.

I. Introduction

Why are there so many different institutions in real world, such as markets, firms, and the government? In markets there are those for final products, for factors, and for intermediate goods. In firms there are those producing goods and services and those marketing them. And in the government there are central and local ones. One of the focuses of economic debate has been that when the market is efficient and when it fails to be so. An answer provided by Arthur Pigou (1920) was that it depends on the existence of external effects. When there is no externality the competitive markets would produce Paretian efficient results. This was later called Welfare Theorems. When there is positive externality, the government should subsidize the agent who produces the beneficial external effects. On the other hand, when there is negative externality, the government should put a *Pigovian tax* on the one who produces harmful external effects. The fallacy of such Pigovian solution was first pointed out by Frank Knight (1924) in a famous example of *two roads*. What Knight emphasized was that the problem proposed by Pigou is not a matter of subsidy or taxation. It is concerned with who has the *property rights* of the roads.

Knight's insight was put forward with a large and important step by Ronald Coase (1960), who found that "...without the establishment of this initial delimitation of rights there can be no market transactions...But the ultimate result (which maximises the value of production) is independent of the legal position if the pricing system is assumed to work without cost." (p. 8) Knight was right but incomplete because it is transaction cost which causes the partial delimitation of property rights and therefore the inadequacy of Pigovian solutions, which implicitly assumed zero transaction cost. And in the real world the transaction cost can not be zero. The question we need to ask is therefore why there are positive transaction costs, and where are they from? What are the effects of these costs on the formation and evolution of various institutions we have observed in the real world?

The logic in Coase (1960) is actually consistent with his famous paper on *The Nature of the Firm* (Coase (1937)). There are transaction costs when either using markets or using firms. The boundary of firms could be determined by equating these two costs at the margin. Actually in the first Section of his 1937 paper, Coase said: "It is hoped to show in the following paper that a definition of a firm may be obtained which is not only realistic..., but is tractable by two of the most powerful instruments of economic analysis developed by Marshall, the idea of the margin and that of substitution, together giving the idea of substitution at the margin." (pp. 386-387). And in another important paper on marginal cost pricing, Coase (1946) had discussed the problem of two-part pricing. In a case of this paper, where the consumer needs travel by some carriage to buy products in a central market, Coase had said: "...the consumer should not only pay the costs of obtaining the additional units of the product at the central market, he should also pay the cost of carriage." What Coase had actually been concerned with was the pricing at two margins, one at the *extensive margin* (the cost of using carriage) and the other at the *intensive margin* (the cost of additional consumption of the product).

In my opinion the big three papers of Professor Ronald Coase (1937, 1946, 1960) can be summarized in one theme: the substitution at multiple margins. Obviously in Coase (1937) there are at least three margins: producing goods by the original firm, producing goods by other firms with lower costs, buying goods directly at the markets if the marketing cost is lower. There are complicated substitution possibilities in such situation, and the equilibrium should be balancing the costs at multiple margins. In Coase (1946), as described above, there are at least two margins to make decisions at. Finally, in Coase (1960), the delimitation of property rights has been concerned with many marginal decisions. In a famous paragraph Coase had said "…the problem is one of choosing the appropriate social arrangement for dealing with the harmful effects. All solutions have costs and there is no reason to suppose that government regulation is called for simply because the problem is not well handled by the market or the firm." (p. 18). This means that we need to compare with each other the costs in using different institutions. This must induce a lot of marginal decisions.

Almost all economic decisions have to be made at multiple margins. This is because scarcity induces competition, and competition would come from many dimensions of real life, such as from those across persons, states, time, locations, and even ideologies and cultures. Competition is just the substitution for something else you want. So by perfect competition we mean perfect substitution at all margins. These margins typically include those of information, goods, and the ease of entry and exit. This is why perfect substitution is usually defined as a situation where all goods are perfect substitutes (homogeneous goods), information is perfect, and there are free entry and exit. This is an ideal world where there is no friction and hence no transaction costs.

Now the question is: why are there transaction costs? The usual answers would be that there are costs involved in bargaining, searching for information, enforcing contracts, and measuring goods, etc. This statement is certainly right. But what we want to argue in the current paper would be that the above costs are indeed the results of non-substitution at some margin(s), and we can use this structure to recover transaction costs or the costs of using institutions. Because transaction costs are so important in determining resources allocation and it is not easy to model them and to measure them in reality, this makes this great idea to be more cited than used. But if we can use substitution structure to trace way back to the determinants of transaction costs, then maybe we can make this idea more tractable. Eventually and hopefully, we would like to see that a theoretical model with Coasian transaction costs could be built and *used* to compare with the world in real life, not just using *blackboard economics* as the only explanation of reality.

In this paper we want to argue that (1) Coase (1937) is actually about the *substitution at multiple margins*, and the resulting *substitution structure*, (2) transaction cost is actually a revelation of the underlying substitution structure of the economy, so if we can describe this structure, then we can understand what the transaction costs would be, (3) perfect competition should be defined as a frictionless world where there are no transaction costs, (4) a world without transaction costs is one in which all goods are perfect substitutes at all margins (or *dimensions*), (5) it is impossible to have complements if transaction cost is zero, and (6) most of the complements in our mind are indeed substitutes *if* some margins ignored are included, so we can not define substitutes/complements only using the usual (Hicks-) Allen elasticity of substitution (AES), which is a single-marginal (or one-dimensional) definition of the substitutability among goods.¹

Section II discusses the substitution structure. Section III uses the carriage example in Coase (1946) and the concept of *dissipation of rents* in Steven Cheung (1974) to illustrate how to use the substitution structure to recover transaction costs. Section IV deals with a real example in Taiwan and shows how the substitution structure works. Section V concludes.

II. The Substitution at Multiple Margins

In neoclassical model, especially that with the assumption of perfect competition, the total revenue is equal to total factor cost, which leaves firms no room for long-run

¹ Blackorby and Russell (1989) had proposed that Morishima elasticity of substitution (MES) is a better representation of the ease of substitution, which happens to be a multiple-marginal (or multi-dimensional) definition of substitution. They also argued that, though AES has also dealt with the case of many goods, it is not an adequate measure of multi-dimensional substitution.

profits. In fact, this assumes that there are no intermediate firms between consumers and producers. No marketing is required. The goods produced by a firm would be directly and instantly delivered to consumers without incurring any costs. Therefore, if the factor cost per unit of goods is \$5 (cost of labor, capital, land...), then the competitive price of this good would also be just \$5, no more and no less. This is impossible. There would be many firms operating in between, which help lower the would-be transaction costs when goods are to be produced and delivered to various consumers. There are only two possibilities for neoclassical competitive market to work. First, the producer is also the consumer such that there is no marketing cost at all, and therefore no need for intermediate firms. Second, there are in-between firms but the costs of using them, the marketing or transaction costs, are assumed to be zero. In this case these intermediate firms are fictitious and have no reasons to exist.

If there are transaction costs in marketing goods produced to consumers, then the reduction of these costs consists of the profit paid for the intermediate firms. For example, if the price of goods consumers finally buy at is \$20 per unit, then this means that the total marketing cost for intermediate firms is just \$15 (\$20 minus \$5), and this is the profit for these firms, which are of course not zero assumed by the neoclassical competitive model.

Although we know that the extra profits come from the reduction of transaction costs, we still need to know what the equilibrium conditions would be, and why there are transaction costs and why in some cases they are larger, and lower in others? Now we use some definitions to organize the thoughts we have had at this moment.

Definition 1: A perfectly competitive equilibrium (C.E. or perfect competition) is one without transaction costs.

Definition 2: In a world without transaction costs all goods are perfect substitutes at all margins.

Definition 3: Any two goods which are perfect substitutes at all margins have an infinite value of the Morishima elasticity of substitution (MES).

These three definitions are indeed a restatement of the arguments in Section I. Under these definitions, we can reach some useful lemmas.²

 $^{^2}$ These lemmas need proof. This would be done in a more complete version of this paper. The current one is only preliminary.

Lemma 1: If there are no transaction costs, then under Definition 1 there could be no distinction between perfect competition and imperfect competition. This is because there could not be imperfect competition if there are no transaction costs. Without transaction costs the only market structure would be that of perfect competition.

Lemma 2: If there is just one margin available, then there would be no complements at all. All goods should be substitutes (not necessarily perfect substitutes, of course) at some margin(s) if there are at least two margins. Most of the cases we have usually considered as complements are actually substitutes if we take more margins, therefore more substitution possibilities, into account.

Lemma 2 is actually an extension of Hicks' definition of *net substitutes*. It was said that if there are only two goods, then they must be net substitutes. His definition should be modified because two goods do not imply one margin. There could be more than one margin in Hicks' two-good case. For example, there are two goods, but with more than one period or more than one state-of-nature. The key point is the number of margins for substitution, not necessarily the number of goods to be substituted for.

Lemma 3: The only possibility for two goods to be complements would be that these two goods do not have any chances to substitute for each other at any single margin. This could be the situation where transaction costs are prohibitively large, probably infinite, such that the barrier of entering any different margin is too high for any single consumer or producer to have incentives to exchange.

Once we have the above definitions and related lemmas, we can use them to construct an *operational definition* of transaction costs. To avoid unnecessary mathematics, in next Section we use some simple cases to illustrate the usefulness of the definitions and lemmas derived so far without building explicit mathematical model.³ One of the cases is from Coase (1946) where consumers face a two-margin decision. This would be the simplest example to explain the substitution at multiple margins, or simply the substitution structure. The other case comes from Cheung's (1974) idea of the dissipation of rents. We will find that there is an identity resulting from the substitution structure which could be used to recover transaction costs through the concept of rents.

³ Actually I am working on a simple model with necessary mathematics such that we can explicitly see what is going on in a model with transaction costs. I always put in mind that Professor Coase does not oppose mathematics. I believe what he would oppose are *unnecessary mathematics* and models which have nothing to do with facts in real life.

III. Examples

Coase (1946) had used an example to criticize the popular Hotelling-Lerner theory of marginal cost pricing. Hotelling-Lerner solution requires that marginal cost be equal to price, or MC = P, and if there are losses due to decreasing average cost, then the government should levy a tax to raise the funds necessary for firms to cover the fixed cost and then to produce the good. This solution is fallacious because users, for example the consumers in the case of carriage, do not pay the full cost which includes the fixed cost of introducing carriages. Consumers should pay for both costs. It is not a matter of taxation.

In an interesting paper Cheung (1974) had analyzed the economic consequences of price control. His analysis showed that there might be dissipation of rents when the government uses price regulation to control commodity prices or the rent of houses. Because of regulation consumers and producers have incentives to change their behavior such that they can do their best to capture these rents. In addition, because of transaction costs the property rights of rents are usually not well defined, the competition for these rents would result in rent dissipation. Some rents would not be obtained by anyone. They just disappear. And the social objective should be to minimize the dissipation of these rents.

Though Cheung provided an interesting view about how transaction cost might work theoretically, he did not derive the equilibrium conditions or constraints which would relate transaction costs to dissipation of rents. This relation might be important because without it we still could not understand why transaction cost is such an important idea in evaluating the performance and evolution of institutions. For example, Gordon Tullock (1967) had initiated the discussion of *rent seeking*, an idea in which there is deadweight loss due to the competition for economic rents. This loss could be a result of corruption, bribery, tariff, etc. But what is the distinction between dissipation of rent and rent seeking?⁴

In my view what distinguishes between these two concepts would be the role played by transaction costs and the substitution structure. As argued by Barzel, rent seeking "neglects the opportunities to gain through avoiding waste" (1997, p. 14). This implies that consumers and producers not only have incentives to waste (through rent seeking) but have motives *not* to waste (through minimizing the dissipation of

⁴ Yoram Barzel (1997, p. 14) had made a provocative distinctiveness as follows: "The "rent-seeking" approach...tends to ignore (almost to a fault) gains from exchange; it concentrates on people's efforts to capture wealth from each other and neglects the opportunities to gain through avoiding waste."

rent). Since there are transaction costs in delimiting property rights, resources owned by nobody become as rents, and to capture these rents would generate transaction costs. If people's incentive is only to seek rents, then the resulting transaction costs for a society would be larger and larger, and no one would gain eventually. So there must be some counter factors which would balance the harmful effects of rent seeking. These factors are just the incentives to minimize the dissipation of rent.

On the other hand, some people would ask why in some cases it seems that only rent-seeking behavior prevails, while in others the rent would dissipate. My answer to this question would be that this is because we usually could not observe all of the relevant margins, at which decisions are made, for the person(s) we want to analyze. Only this person knows how many margins she should take into account. Other people could only judge her behavior from her decisions. And we just could not obtain all useful information due to transaction costs in searching for information. But in any case, from the discussion of Cheung's idea of the dissipation of rent, we still could have derived an identity which must hold in any equilibrium with positive transaction costs. This identity would relate dissipation of rent to transaction costs.

Identity:
$$\sum_{i=1}^{N} (\Delta C_i^T + \Delta R_i^T) = 0$$
, with $C_i^T \ge 0, \forall i$.

This identity includes both the example of carriage in Coase (1946) and that of price control in Cheung (1974) as special cases. In this identity, ΔC_i^T indicates the change of transaction costs (C^T) in the *i*th margin, ΔR_i^T is the change of rents (R^T) in the *i*th margin, and *N* is the maximum number of relevant margins people could take into account when they make decisions, but usually they could not consider all of them due to transaction costs (costs of searching for full information). The following are some special cases of this identity, which are useful in later discussion.

Example 1 (Perfect Competition): Since in perfect competition or in a competitive equilibrium there are no transaction costs ($C^T = 0$), this implies that the change of transaction costs would also be zero. On the other hand, the rent is defined as the product of the difference between price and marginal cost and the quantity produced ($R^T = (P - MC)Q^*$). Then by the First Welfare Theorem we know that a competitive equilibrium is Pareto optimum, and this implies that price would be equal to marginal cost, which squeezes the rent down to zero ($R^T = 0$). And this in turn implies the

change of rent is also zero, so the Identity above would be trivially satisfied.

Example 2 (Monopoly): In the other extreme of market structure is a single firm which acts as a monopoly at all margins. By "at all margins" I mean that there is no way for the monopoly status of this firm to dissipate. In this case this monopolist must do all things she can to prevent the invasion from any other firms in any possible means. This could generate huge transaction costs such as legal restrictions, government regulation, expenditures for protection from outside competitors, etc. The monopoly rent she gets at any margin, say barrier to entry by a legal restriction, would be at the cost of consumers and other non-incumbent firms at some other margin(s). These costs are transaction costs because without the legal restriction a market would work. The transaction costs would not necessarily occur at the same margin as that restrained by the law. But they never disappear. They just hide somewhere or, in my words, at some other margin(s). The amount of the rent obtained by the monopolist would exactly equal to the amount of the resulting transaction costs if we sum over all available margins. This is what the above identity has told us. That is, the sum of (i) increase of rent of the monopolist, and (ii) dissipation of rents from consumers and

other potential, non-incumbent firms, could be denoted by: $\sum_{i=1}^{N} (-\Delta R_i^T)$, the net

decrease of rents summed over all consumers and producers at all relevant margins. This must be equal to sum of the net increase of the transaction costs at all margins: $\sum_{i=1}^{N} \Delta C_i^T$. The equality of these two sums (i.e. $\sum_{i=1}^{N} (-\Delta R_i^T) = \sum_{i=1}^{N} \Delta C_i^T$) is just a restatement of the Identity.

An example came from the famous *conjecture* of Coase (1972), where a durable good monopolist would not hold her status of monopoly if she could not make credible commitment on the amount of land (or natural resources) she would like to develop (or produce) in the future. This monopolist would finally act just like a firm in perfect competition. The only difference would be that there is still only a single firm in operation, but the monopolist will compete with her future-self. In other words, there is an *intertemporal margin* out there where the monopolist must decide whether to keep promise and develop no more land in the future, or to break commitment and develop land further as more as it is available.

In the case that there is credible commitment for the monopolist, the incumbent rent of consumers (or land buyers) would not dissipate, and therefore transaction costs would not increase further. But if the monopolist could not make credible commitment, then she could make more profits, or monopoly rent, at the cost of incumbent buyers of land, whose rents would therefore dissipate. And in the meanwhile, there are transaction costs out there, such as costs of collecting information about the credibility of monopolist's commitment, costs incurred when competing with other buyers to sell the land in hand because the price of it would decrease due to the sale of more amount of land by the monopolist, etc. In each of these two cases the Identity would be satisfied because in the former case (with commitment) it is just like the monopolist case in Example 2, and in the latter (without commitment) it is very similar to the perfect competition case in Example 1.

Example 3 (Dissipation of Rent): In Proposition 2 of Cheung (1974) he had argued that "Given the existence of non-exclusive income and its tendency to dissipate, each and every party involved will seek to minimize the dissipation subject to constraints. This will be done either through seeking alternatives in using or producing the good so that the decline in resource value is the lowest, or through forming alternative contractual arrangements to govern the use or production of the good with the least rise in transaction costs, or through the least costly combination of the two procedures." (p. 61). What Cheung wanted to say was that consumers and producers would always attempt to minimize the dissipation of rent subject to constraints. I think we can rephrase Cheung's proposition as the following minimization problem:

Program 1 (Individuals' Optimization Problem (Resulting in New Institutions)):

Minimize
$$\sum_{i=1}^{M} (-\Delta R_i^T)$$
, subject to $\sum_{i=1}^{N} (\Delta C_i^T + \Delta R_i^T) = 0$ and $C_i^T \ge 0, \forall i$.

The dissipation of rent in the *i*th margin is simply $-\Delta R_i^T$. Since Cheung had emphasized that individuals would seek all possible alternatives to minimize the dissipation, this means that they will try to minimize dissipation at relevant margins (to use the phrase in this paper), and hence to minimize $\sum_{i=1}^{M} (-\Delta R_i^T)$.⁵ According to the Identity, minimizing the dissipation of rent is equivalent to minimizing the rise in transaction costs. This is obviously consistent with what Cheung had phrased in his Proposition 2.

⁵ The *constraints* mentioned in that Proposition would be the Identity in the current paper, though I have not confirmed this with Professor Cheung. Note that the number of relevant margins individuals would take into account (M) is assumed to be no more than that available in maximum (N).

The difference between this paper and Cheung (1974) would be that we do not think that individuals would take all the harmful effects they produce on others into account. They could only take *some* of them into account ($M \le N$). This behavior is still consistent with the general philosophy of minimizing the dissipation of rent. Why? The answer to this question is simple: an individual's action with harmful effects on the third party would possibly increase her benefit or rent, and this makes sense when there are transaction costs because in such a situation some of the rent would be non-exclusive, that is the property rights of goods would not be well delimited. Only when this negative externality does harm to individual who generates it would this person try to avoid further harmful effects, but only for her own benefits. There is no *a priori* reason to assume that people would internalize (not necessarily *intentional*) all harmful effects by minimizing the dissipation of rent. Sometimes they would, but definitely not for all times if there are transaction costs, which would make the internalization of external effects impossible and inefficient. From the above discussion we have the following definition for a social optimum (not necessarily in the sense of Pareto):

Program 2 (Social Optimization Problem (for a Fictitious Planner)):

Minimize
$$\sum_{i=1}^{N} (-\Delta R_i^T)$$
, subject to $\sum_{i=1}^{N} (\Delta C_i^T + \Delta R_i^T) = 0$ and $C_i^T \ge 0, \forall i$.

Note that the difference between Program 1 and Program 2 is that individuals would not take all margins into account when they try to minimize the dissipation of rent, but a fictitious social planner would. What makes this comparison any different from the usual Welfare Theorems? To put it simply, the Welfare Theorems in the Pigovian tradition implicitly assume that there are no transaction costs in the use of any institutions including markets, firms, and the government. This is obviously wrong!⁶ But the creators of Welfare Theorems in Pigovian tradition did not tell us what will be going on if there *do* have transaction costs.⁷ Here we provide such an example where there are positive transaction costs. When transaction costs are assumed zero, as did implicitly by Pigou and explicitly by Arrow and Debreu, the change of transaction costs would be trivially zero, and if *C.E.* = *P.O.* prevails, then there would be no rents (for *P* = *MC*) and hence no dissipation of them. The objective functions in both

⁶ We owe this rejection to Pigovian tradition to Knight (1924) and Coase (1960), though Professor Knight did not mention the idea of transaction costs. He had only talked about property rights.

⁷ The two Welfare Theorems combined give the result of C.E. = P.O., that is competitive equilibrium is equivalent to Pareto optimum. As is well known, this result was discovered independently by Professors Kenneth Arrow and Gerard Debreu in 1951.

Programs 1 and 2 would have values equal to zero. So these two Programs are equivalent no matter what the values of the numbers of margins M and N are. This is equivalent to the implication of Welfare Theorems.

But when transaction costs are positive, the objective functions in Programs 1 and 2 diverge. Because $M \le N$, the value of the objective function in Program 2 is almost always smaller than that in Program 1. This is the usual meaning of a social optimum. It is by definition better than that induced by individuals. And if there are positive transaction costs, then C.E. and P.O. are both unreachable. They are ideal worlds to which we do not belong! We live in a real world where there are positive transaction costs. In such a world the only chance for individuals' equilibrium to coincide with social optimum would be in the case of M = N,⁸ that is every individual faces the same substitution structure as the fictitious social planner does, such that they would solve the same problem at the same margins. This is quite demanding because a single individual or even a cohort of individuals usually could not have so much information as the fictitious planner could. So from the social point of view a good government policy should be to lower the barriers to entering more extensive margins as possible as it can. This is equivalent to lowering the transaction costs in the use and formation of institutions, which might in turn foster more and more better institutions.

Example 4 (Marginal Cost Controversy): Coase (1946) had proposed two-part pricing as an alternative to Hotelling-Lerner marginal cost pricing solution when the average cost is decreasing. Transaction costs had not been explicitly discussed in that paper. When firms maximize profits they set marginal cost to marginal revenue (MR = MC), and there is a monopoly rent because price is higher than marginal cost (P > MC). Under two-part pricing, marginal cost is still equal to price in the second part of pricing (or at the *intensive margin*), but consumers should pay for the fixed cost of the use of carriages $((AC - MC)Q^*)$ in the first part (or at the *extensive margin*). If we compare the MR = MC scheme with that of P = MC in the second-part pricing, then we should find that in the first scheme there is a monopoly rent, but in the second the rent would be zero. There is dissipation of rent! Will the amount of this dissipation just be equal to the fixed cost of using carriages? No, it is not necessarily so. This is because the fixed cost may not necessarily include all the expenses needed to bring carriages into existence. For example, if the cost of carriages is too high such that no firms and consumers could afford it, it might be provided by the government as a public good. But using government has its cost, a kind of transaction costs. In this

⁸ The individuals' equilibrium could be any kind of institutions such as market, firm, and even the government.

case this extra transaction cost should be included as part of total cost. But in any case our Identity would still hold. The total dissipation of rent would always be equal to the total increase in transaction costs. If we look at only one margin (number of shopping trips taken), and ignore others (whether there is a carriage, or whether the carriage is provided by the government through taxation or by consumers who pay for themselves), then we might reach a wrong conclusion such as that suggested by Harold Hotelling and Abba Lerner.

IV. How the Substitution Structure Works: A Case Study in Taiwan

Cheung (1983) had emphasized that what Coase (1937) was really concerned with is the contract. This is why Cheung's paper was titled "The Contractual Nature of the Firm". He claimed that the distinction between market and firm is vague because "The word "firm" is simply a shorthand description of a way to organize activities under contractual arrangements that differ from those of ordinary product markets." and then he said "The growth of a firm may then be viewed as the replacement of a product market by a factor market, resulting in saving in transaction costs...Coase does not define "the firm"; nor...is there a clear distinction between a product market and a factor market." (p. 3). The point of Cheung, if not misunderstood, was that what consumers and producers choose are contracts, so different contractual arrangements would result in different institutions. The nature of the firm is therefore a problem of the substitution of contractual arrangements. Though Cheung's emphasis on contracts is definitely right, his view of firm "as the replacement of a product market by a factor market" needs some modification. After all a factor market is still a market, not a firm.⁹ And not all activities could be written in explicit contracts. What if there are implicit contracts such as those related to tradition, custom, and even ideology? Though many of these contracts would not be measurable, they could still induce transaction costs and affect the performance of the institution. Perhaps we need a more fundamental concept to deal with both explicit and implicit contracts.

The substitution structure proposed here in this paper is doing this thing. No matter it is the substitution of contractual arrangements as argued by Cheung, or the conclusion proposed by Coase that "At the margin, the costs of organizing within the firm will be equal either to the costs of organizing in another firm or to the costs involved in leaving the transaction to be "organized" by the price mechanism." (1937,

⁹ More precisely, in Cheung's mind it should be that a firm can be viewed as "the replacement of the contractual arrangements in a product market by those in a factor market".

p. 404), what in common is that they are both concerned with the substitution at multiple margins. In Coase (1937) there are at least three margins: organizing within the firm, organizing in another firm, and organized by the price mechanism. And in Cheung (1983) any two of the contracts would be a margin so there could be many margins at which to make decisions. The performance of an institution is determined by its substitution structure, and the evolution of institutions would in turn determine the substitution structure of a society. In the following we would use the raise in Taiwan's minimum wage as an example to see how the substitution structure works.¹⁰

The hourly minimum wage in Taiwan was raised from \$98 to \$103 (in New Taiwan Dollar) in January 2012, and the monthly one also rose from \$17,880 to \$18,780 at the same time. There had been debates on whether the unemployment rate would increase due to the rise in minimum wage. Now we would see what will happen from the perspective of substitution structure. There are at least three margins for firms to get rid of the rise in wage costs. First, if the rise in minimum wage is small enough, firms can slightly lower the amount of product, say packing ice tea in a smaller bottle; or they can slightly reduce the quality of product, say baking a cheese cake with less cheese. This behavior would increase the rent of firms but at the same time would increase transaction costs of consumers (because they need to visit more stores to find products with a higher quality), or would cause the rents of consumers to dissipate.

Second, if the rise in minimum wage is small but larger than that in the first case, especially as in Taiwan the hourly wage had increased at a higher rate than that in monthly wage, then firms would lay off some of the part-time workers or reduce the hours worked for those with a full-time job. So the relevant margin to make decisions is now at the one of working hours rather than at that of the quality of product, as in the first case. Finally, if the rise in minimum wage is large enough, especially for the monthly wage, then firms would be forced to fire some of the full-time workers to reduce their labor costs. In Taiwan the last case has not happened because the rise in monthly minimum wage is less than 10%, which is not big enough as a reason to fire people, because layoff would itself have costs.

These three margins in the case of Taiwan's minimum wage would have

¹⁰ Actually I am currently working on such a model which could be applied to the discussion of economic growth and development. I believe that nothing can be more important in fostering growth than policies that could effectively reduce transaction costs. This is because, according to our theory, a decrease in transaction costs is equivalent to an increase in rents (or a decrease in the dissipation of rents). Total product of a society would be maximized if the total rents are maximized.

interesting implications for the substitution structure. For example, because firms have three margins to make decisions at, they must compare the costs of substitution with each other when using different margins.¹¹ When the rise in wage is smaller, firms at the margin of the quality of product would most easily substitute the rise in transaction costs borne by consumers, who would spend more time in searching for high-quality product, for the increase in labor cost they have to pay. But either at the margin of hours worked or at the margin of number of persons to hire, firms have to incur more transaction costs if either they hire less or employees work less. Firms would therefore behave just like minimizing transaction costs, or in our terms like using the margin at which they have the best chance to substitute low-cost strategy for high-cost one. This chance of substitution would be determined by how many margins there are for firms to choose at and what the transaction costs at these margins are. The easier for firms to substitute for high-cost strategy the less the transaction costs incurred would be. And because there would be many relevant margins, we could not just look at one single margin to determine which is the cheapest way to do things right. We must put all relevant margins together and then find the least costly way to make decisions. This is what we mean by substitution at multiple margins.

V. Conclusions

In this paper we have constructed a simple theoretical framework to discuss how Coase's great idea of transaction costs could be incorporated into a simple model and how it works. We consider transaction costs as a revelation of the substitution structure of an institution. Consumers and producers would try to minimize their costs when making decisions. The margins at which decisions are made are usually multiple, and economic agents would use the least costly way to achieve their goals. The substitution structure of an institution reflects possibilities for agents to substitute cheaper means for expensive ones. Because there are multiple margins, the cheapest way can only be found by searching for all relevant margins. In doing so economic agents are trying to minimize transaction costs or, in terms of Cheung (1974), to minimize the dissipation of rent.

In a paper I am now working on, the ideas in this paper would be extended to the study of economic growth and development. Transaction costs would be determined by the substitution structure or the ease for economic agents to substitute at multiple margins. Morishima elasticity of substitution (MES) would be used as a measure of

¹¹ Weizsacker (1984) had coined the term "costs of substitution", but he had a different interpretation.

this ease of substitution because it is a natural definition of multi-dimensional substitution.¹² Since transaction costs are usually not easily measured, we need to find proxy variables such that we can still evaluate their effects through measuring the proxy. One of the contributions of the present paper would be that we can use rents as a proxy to measure transaction costs. This is because we have derived an Identity which relates the rise in transaction costs to the dissipation of rents. To sum over all relevant margins the total increase in transaction costs must be equal to the total dissipation of rents in an economy. And because it is easier to measure rents than transaction costs, this makes empirical works much easier to undertake. We hope there will be more empirical works to be done in this direction. Then we can see how institutions would evolve and how they would be shaped by the activities of consumers, producers, and the government.

References

- Barzel, Yoram (1997): *Economic Analysis of Property Rights*, 2nd ed., Cambridge, England: Cambridge University Press.
- Blackorby, Charles and R. Robert Russell (1989): "Will the Real Elasticity of Substitution Please Stand Up? (A Comparison of the Allen/Uzawa and Morishima Elasticities)," *American Economic Review*, 79, 882-888.
- Cheung, Steven N. S. (1974): "A Theory of Price Control," *Journal of Law and Economics*, 17, 53-72.
- ---- (1983): "The Contractual Nature of the Firm," *Journal of Law and Economics*, 26, 1-22.
- Coase, Ronald H. (1937): "The Nature of the Firm," Economica, n.s., 4, 386-405.
- ---- (1946): "The Marginal Cost Controversy," Economica, n.s., 13, 169-182.
- ---- (1960): "The Problem of Social Cost," Journal of Law and Economics, 3, 1-44.
- ---- (1972): "Durability and Monopoly," Journal of Law and Economics, 15, 143-149.

Knight, Frank H. (1924): "Some Fallacies in the Interpretation of Social Cost," *Quarterly Journal of Economics*, 38, 582-606.

- Pigou, Arthur C. (1920): The Economics of Welfare, London: Macmillan.
- Tullock, Gordon (1967): "The Welfare Cost of Tariffs, Monopolies and Theft," *Western Economic Journal*, 5, 224-232.
- Weizsacker, C. Christian von (1984): "The Costs of Substitution," *Econometrica*, 52, 1085-1116.

¹² Since the discussion of the measure and application of MES is more mathematical and has to deal with a formal model, it needs another paper to address this issue and is therefore omitted here.