IMPERFECT COMPETITION IN MODELS OF WAGE FORMATION AND INTERNATIONAL TRADE

av

HENRIK HORN AF RANTZIEN

AKADEMISK AVHANDLING

som för avläggande av ekonomie doktorsexamen vid Handelshögskolan i Stockholm framlägges till offentlig granskning måndagen den 10 oktober 1983 kl 10:15 i sal TORSTEN å högskolan, Sveavägen 65.

Stockholm 1983

IMPERFECT COMPETITION IN MODELS OF WAGE FORMATION AND INTERNATIONAL TRADE

A Dissertation for the Doctor's Degree in Economics
Stockholm School of Economics, 1983

ISBN 91-7146-415-8
IDK 339, 137, 331, 2, 339, 5

Henrik Horn af Rantzien

IMPERFECT COMPETITION IN MODELS OF WAGE FORMATION AND INTERNATIONAL TRADE

INSTITUTE FOR INTERNATIONAL ECONOMIC STUDIES
University of Stockholm

PREFACE

When I started to work on the present study, I was firmly determined to write an empirical dissertation. My original intention was to apply the so-called product cycle theory to some Swedish exportables. But the more I studied this theory, the more unclear it appeared to me. Instead I became increasingly interested in one of its fundaments - imperfect competition - and ended up writing an entirely theoretical thesis on issues with only a very weak, if any, connection to the product cycle theory.

Many people contributed in different capacities. My first thought goes to my family and to my friends Peter Brundell, Lisa Adebo, Mia Lindqvist, Per Hedberg and Per Wärnegård. I am truly grateful to them for their patience with me.

I was lucky enough to have the opportunity to carry out the work at the Institute for International Economic Studies at the University of Stockholm - an advantage that can hardly be overemphasized. The Institute offers an environment that is ideally suited for research with its competent research and secreterial staff and prominent visitors. There are in particular two collegues to whom I am indebted: Lars Svensson and Torsten Persson. Lars was my thesis adviser. His efforts, however, went far beyond what a student can expect. Lars was always ready to read and discuss my attempts at formulating and solving theoretical problems, always willing and able to explain what I did not understand. Torsten, with his creative mind and positive disposition performed much the same task. It may seem as a cliché that without their guidance and encouragement the work would not have been completed, but it is nevertheless not far from the truth.

I also benefitted from valuable help from other collegues at the Institute. Carl Hamilton and Assar Lindbeck, in particular, provided guidance and much-needed moral support, as did Refik Erzan, Harry Flam and Peter Svedberg.

The international contacts of the Institute were of major importance to the study. Avinash Dixit, Princeton University, was an inspiring teacher during my stay at the University of Warwick, where he held a chair at the time; Elhanan Helpman, Tel-Aviv University, stimulated my interest in imperfect competition in international trade, as did Wilfred Ethier, University of Pennsylvania, and Henryk Kierzkowski, Graduate Institute of International Studies, Geneva; discussions with Gene Grossman, Princeton University, influenced the chapters on wage formation; Assaf Razin, Tel-Aviv University, at an early stage made a specific suggestion which eventually became chapter two and thereby got me started. I am grateful to them all for their intellectual stimulus and moral support.

The secretaries at the Institute minimized my practical problems with skill and patience: Birgitta Eliason provided invaluable help in administrative matters; Sholeh Blom efficiently (and several times) typed major parts of the "final" version of the manuscript; Caroline Burton, Edda Liljenroth, Anita Oxenham, Brita Slotorup and Ann-Marie Soler assisted in typing or in editorial matters.

Jörgen Appelgren and Anders Forslund helped me by checking some of the calculations. It was with mixed feelings I saw the many errors they found. I am furthermore grateful to David Brownstone, Ragnar Lindgren and Mats Persson for accepting the tedious task of being discussants of various chapters in seminars at the Stockholm School of Economics.

Finally, financial support from Jacob Wallenbergs Forskningsstiftelse, Humanistisk-Samhällsvetenskapliga Forskningsrådet, and Riksbankens Jubileumsfond is gratefully acknowledged.

CONTENTS

CHAPTER I

INTRODUCTION

CHAPTER I INTRODUCTION

Perfect competition is by far the most thoroughly analyzed market structure in post-war economic theory. Adherents to the model of perfect competition justify its use by arguing either that its underlying assumptions are resonable and analytically convenient approximations of the real world, or that the model explains real world events well. These arguments are, however, not valid in many important instances.

First, most economists agree that the basic assumptions underlying the perfect competition model are often questionable on empircal grounds. Products are often differentiated, rather than homogeneous; cars, stereo equipments, and cameras are typical cases. Fixed costs, due to e.g. R and D or distribution investments, give rise to falling average costs. Some markets are dominated by one or few firms and these firms do certainly not take the price of their products as given. Furthermore, many factor markets also seem to fit poorly into the perfect competition paradigm. Markets for two important factors - labor and oil - are clearly characterized by strong non-competitive features. The oil market has the last decade been strongly influenced by a producer cartel and a small number of oligopolistic retailers. The labor market has, in particular in many European countries, been characterized by collective bargaining between large employers (federations) and trade unions, rather than by decentralized wage setting.

Second, it is easy to find many contemporary real world economic problems on which analyses within the perfect competition framework have thrown little light. For instance, international trade has increased dramatically since World War II, and in particular has intra-industry trade expanded. It is estimated that around fifty percent of world trade is of the intra-industry type.[1] It has not been possible to construct models that are based on the basic assumptions of traditional trade theory, (i.e. on homogeneous products, constant returns to scale, and perfect competition), that are able to generate this type of cross-hauling of similar products. Another example is provided by one of the major macroeconomic issues during the last decade - the unemployment problem. Its origins have probably differed to some extent across countries, but it seems that too high real wages are at least part of the problem in the smaller European market economies. But despite persistent unemployment, real wages have failed to adjust sufficiently in these economies.

Since these issues (and others) are not understandable within the theory of perfect competition, it seems required to base the search for clues to these phenomena on other premises about market structure and individual agents' behavior. This is indeed the strategy of the present study.

The following five chapters are all attempts to crossbreed different fields in economics. The important unifying theme is that in all chapters elements of imperfect competition are introduced into fields where perfect competition has been the predominantly assumed market structure. Chapters II, III and IV deal with imperfect competition in the goods market. The type of model employed in these chapters is inspired by the literature on industrial organization, and it has recently been employed in international trade theory for analyses of intra-industry trade. Chapters V and VI study imperfect competition in the labor market, in the form of a monopolistic trade union. The issues studied in these chapters are usually found under the heading of macroeconomics, but the methods stem partly from labor economics.

The following sections give a brief background to, and a selective presentation of some approaches that have been employed in the previous literature to model imperfect competition. Sections 2 and 3 also contain short presentations of chapters II through IV, and V plus VI, respectively. Section 4, finally, gives some concluding remarks.

1. IMPERFECT COMPETITION IN ECONOMIC THEORY

The awareness of imperfect competition is very old indeed; it was Aristotle (341 B.C.) who coined the expression "monopoly".[3] But even though the phenomenon was well-known, and utterly disliked, it was nevertheless not subject to much stringent analysis by economists, until Cournot (1960) in 1838 published his path-breaking analysis. Marshall (1920), the patron saint of the theory of *monopolistic competition*, who's emphasis on marginal analysis was inspired by Cournot, developed further the analysis in a partial equilibrium framework.[4]

The foundation of the present theory of monopolistic competition was laid early in the 1930s by Chamberlin (1933) in *The Theory of Monopolistic Competition*, and by Robinson (1933) in *The Economics* *of Imperfect Competition*. These studies ended a debate, known as the *Economic Journal Controversy*, that had started with an article by Sraffa (1926). He claimed that firms cannot have decreasing average cost in a perfectly competitive equilibrium, since this would run counter to the assumption of profit maximization. Several well-known economists contributed to the debate, including e.g. Harrod, Pigou, Robbins and Schumpeter. The concept introduced by Chamberlin (1933) and Robinson (1933), that managed to end this debate by combining internal increasing returns to scale with competition, was that of *product differentiation*.

The theory of monopolistic competition was anticipated by many to revolutionize economic analysis, but it did not live up to these expectations. The literature after World War II on imperfect competition was instead mainly concerned with markets with few firms, and with game theory.[5] But the market structure analyzed in this literature was mainly perfect competition. General equilibrium analysis, founded by in particular Walras (1954) early in the 1900s, and cast into a more rigorous framework by Arrow and Debreu (1954), took its starting point in this analytically more straightforward framework. By the early 70s this analysis had reached a point of considerable mathematical complexity. The applicability of the derived results was, however, often less impressive than the complexity of the methods used.

Professional discontent with the perfectly competitive framework increased during the 70s. In its defense it was argued that under certain conditions, the model could describe the economy to which an imperfectly competitive economy converges, when the number of firms becomes large. But it was also clear from a theoretical perspective that perfect competition at the very best could be a crude approximation to normal real world market structures. This view was based on developments in the theory of information, summarized by Salop (1976) as:

> The relevant market structure with imperfect information is not perfect competition but rather monopolistic competition. (p. 240).

As a consequence, interest in alternative approaches surged forward in the second half of the 70s.[6]

Imperfect competition had not been completely abandoned during the heyday of the Walrasian model, particularly not in fields with more emphasis on empirical work. Two such fields of particular relevance to this study are industrial organization, and labor economics. In contrast to the Walrasian general equilibrium approach, the analyses in these fields were almost invariably performed within a partial equilibrium framework. Furthermore, the Walrasian assumption of perfect competition was completely alien to the spirit of the literature on industrial organization and e.g. the literature on trade unions. In these fields the Walrasian model never made any real impact - their prime *raison d'être* was precisely imperfect competition.

2. MONOPOLISTIC COMPETITION IN TRADE THEORY

In search of analytical models that are able to generate the empirically well-established phenomenon of intra-industry trade, trade theorists have turned to the theory of industrial organization. In particular, attention has been directed toward the models of monopolistic competition in differentiated products that were originally developed by Chamberlain (1933) and Robinson (1933).

This theory of monopolistic competition rests on two main assumptions: first, every firm has internal increasing returns to scale, and second, each producer's product has no perfect, but many imperfect, substitutes from the consumer's viewpoint. As a consequence of the first assumption, only a limited number of differentiated products are produced in equilibrium. Due to the second assumption, every producer has some degree of monopoly power, but he also faces some competition from producers of imperfect substitutes.

The recent theoretical literature on intra-industry trade has used a couple of different models to represent product differentiation. That is, different assumptions have been made concerning properties of the space in which products are located. These models have recently been surveyed by Helpman (1983), and the presentation of them here is therefore very brief.

2.1 *Models of differentiated products*

Recent intra-industry trade models fall into either of two principal groups, according to the assumptions made about the product space. The two categories are similar in that there is demand for all products in the respective product space, but they differ in how these demand functions are derived.

One group, using an approach developed by Spence (1976), and Dixit and Stiglitz (1977), takes as its starting point that

> [t]he (strict) quasi-concavity of the utility function means that the consumer desires to consume a *variety* of commodities rather than to consume any one commodity.
>
> (Takayama (1974, p. 150)).

It assumes the existence of a representative consumer whose preferences are represented by a Constant Elasticity of Substitution (CES) utility function. This function has the property that each product is an equally good or bad substitute for every other product, and it usually gives rise to positive demands for a large number of products.

The other group of models, on the contrary, assumes that every consumer prefers one particular good, and that he purchases one product only - the one that is the best substitute for the most-preferred variety. Consumers' tastes are represented by a one-dimensional geometric device, along which all products are ordered. One point corresponds to the consumer's most-preferred specification, and the further away from this point a product is located, the less the utility derived from consuming it. But individuals differ in their most-preferred variety, and with consumers distributed continuously along the product spectrum there will be demand for all products in the aggregate.

Two different representations have been used in the latter approach[7]. Lancaster (1979, 1980) assumes that the product space lacks end-points; an infinite line or the circumference of a circle describes this type of product space. Eaton and Kierzkowski (1983), on the other hand, instead use a finite line. This difference has very strong impacts. With the finite line approach, firms located at both end-points of the line face competition from one direction only, and are thus more protected than other firms. But this in turn has ramifications for adjacent firms, and hence the location of a firm will influence its choice of quantity to produce or price to charge. This decision is immaterial in the Lancaster framework, under the assumption about a uniform distribution of consumers, since it lacks these end-points.

Chapters II through IV all employ the Spence-Dixit-Stiglitz formulation of the demand side. It is definitely not claimed that this is the empirically most attractive approach. Its main advantage lies in its analytical simplicity, but there are many aspects of product differentiation that are essential to a large class of problems, that it does not capture. However, the approach has two desired basic features in that it gives rise to demand functions in the aggregate for a variety of products, and it implies that there is some utility gain involved in increased product variety.[8] The questions considered in chapters II and IV - non-homothetic production functions, and customs unions, respectively - are themselves so involved that additional complexities must be kept to a minimum, at least as a first step. In chapter III the least attractive feature of the Spence-Dixit-Stiglitz representation - the parametrically given elasticity of demand - is circumvented, as explained below.

2.2 *Contents of chapters II, III, and IV*

In the real world products do not usually become differentiated without resources being sacrificed somewhere in the production or distribution process. Typical for differentiated products is that e.g. R and D and/or marketing costs stand for a substantial part of total costs. But there are obviously also other activities performed in the firm, such as the actual production of the physical good. There is no reason why the firm should employ inputs in the same proportion in these other activities, and in the "differentiation" part of the process, nor is there any reason why it should expand both activities proportionally, as it expands output. Production processes of differentiated goods are therefore likely to be non-homothetic, i.e. factor proportions in production of differentiated goods depend on output levels, and not just on relative factor prices.

This observation is the starting point of chapter II. The chapter considers properties of a generalized version of the traditional two-sector, two-factor model underlying e.g. the Heckscher-Ohlin theory. Different versions of the model have recently been employed to study intra-industry trade. One sector is of the traditional type, whereas in the other sector firms exhibit internal increasing returns to scale, products are differentiated, and Chamberlinian monopolistic competition prevails. The chapter's main focus is on how non-homotheticity in this sector's technology affects the dual equilibrium relationships between factor and commodity prices on the one hand, and between factor endowments and industry outputs on the other. For instance, conditions are given under which a Rybczynski-type proposition is valid, and under which autarky values of various variables serve as predictors of intra-industry trade patterns.

Chapter III deals with two questions. The first has received a lot of attention both among economists and in the popular debate: are there too many or too few different variants of the same good produced in a market equilibrium, compared to what would be socially optimal? A popular established view is that there is excess product diversity, even though economists have not been able to form any consensus for or against this view. The second question considered is whether regulation of every firm's output volume (leaving entry and exit decisions free), or regulation of the number of firms (leaving output decisions free), is the most welfare improving second-best policy for cases where the number of firms in the market equilibrium is not socially optimal. It is assumed that the regulation in itself is costless, and that information is perfect. The chapter investigates to what extent the answers to these two questions are affected by the absolute size of the economy, as measured by its factor endowment (number of workers). Changes in the size of the factor endowment can either be interpreted at the result of growth, or the opening up of international trade.

The economy under consideration is similar to the one studied in chapter II, except that there is just one factor of production. It is assumed that the number of firms n is not generally large enough to justify the approximation that $1/n = 0$. The elasticity of demand facing an individual producer in equilibrium is therefore not constant, contrary to what is usually assumed within the Spence-Dixit-Stiglitz framework. Instead the elasticity increases in the number of firms, and as a consequence the nature of the equilibrium solution depends on the number of firms, and hence on the size of the total economy. It is shown that "excess" diversity in the market equilibrium is more likely the smaller the economy is, and that for sufficiently large economies there will be "too little" diversity, relative to our welfare criterion. Furthermore, it turns out that in the very large economy regulation of the number of firms improves welfare more than controlling their output levels.

Customs unions, the topic of chapter IV, have attracted the interest of many economists since the pioneering analysis by Viner

(1950). There is a large body of theoretical analyses of their effects on resource allocation and welfare. The lion's share of this literature is concerned with economies in which a fixed number of goods (usually two or three) are produced under constant returns to scale, and traded on perfectly competitive markets. As argued above, these assumptions are not valid for a large share of goods produced today, nor is such a characterization consistent with a substantial part of the present trade flows. The purpose of chapter IV is to take a step in the direction of analyzing how customs unions affect resource allocation and welfare in economies with product differentiation and imperfect competition. In particular, the chapter focuses on issues that are not present in traditional analyses, such as how customs unions affect the degree of product differentiation.

An analysis of customs unions requires at least three countries. In chapter IV it is assumed that two similar economies with some developed country characteristics have formed a customs union, and hence levy the same tariff on a third country, which is taken to be a less developed country. The DCs both produce differentiated goods, and a homogeneous good, whereas the LDC produces the latter only. It is shown among other things that if the member countries jointly increase their common external tariff the number of produced varieties will be reduced proportionally in both the DCs. Furthermore, if both the DCs introduce a tariff on the partner's differentiated goods, there will be a relative reallocation of production of differentiated goods from the smaller to the larger partner. It is even possible that the larger partner's output of these goods rises.

3. TRADE UNION MODELS

The possibility for an owner of a factor of production to raise its remuneration by a (threatened) check to its supply, was pointed out by Marshall (1920). He gave four conditions for this to be viable, part of the first being that

> .... no good substitute being available at a moderate price. (p. 319).

Few individual workers fulfill this requirement if they are not in possession of some unique skill. The obstacle to the individual worker acquiring a higher wage is precisely that he is normally easily substituted for by other workers. But if other workers refuse to take his job, our worker comes into a position that fits Marshall's (1920) first condition. By coordinating their actions, workers may acquire the power to reduce or eliminate the employer's possibility of profitably maintaining the production in the case of wage disputes, and hence gain increased bargaining power. The vehicle for such coordination is, of course, the *trade union*.

Trade unions played a very modest role in the economic literature preceding World War II, even though e.g. Mill (1909) had already argued in 1848 that unions are a normal part of an economy[9]. Even though imperfect competition among sellers in product markets was a well-recognized and intensively studied phenomenon in the economic literature of the 1930s, the seller's side of the labor market was still mainly viewed as perfectly competitive. However, real world developments made it increasingly clear that an economic theory of trade union behavior was needed. This gave rise to a debate, sometimes referred to as the "Ross-Dunlop Controversy" after its two main protagonists.

3.1 *The Ross-Dunlop Controversy*

The Ross-Dunlop Controversy dealt with whether or not any meaningful analysis of trade unions could be performed within the realm of economic theory. Dunlop (1944) argued that[10]

> [j]ust as the process of product-price formation has been illuminated by economic analysis .... so organized labor markets represent a field of important theoretical inquiry. (p. 4).

This was not in accordance with the views held by Ross (1948), who thought that

> .... many of the most interesting questions concerning union behavior cannot be answered by any strictly economic analysis... (p. 4),

and

> [a]mong all the participants in economic life, the trade union is probably least suited to purely economic analysis. (p. 7).

Ross claimed that trade unions are political organizations, and as such are too complex for traditional economic analysis:

> [w]hat is needed is to break down the walls between the separate disciplines of social science which have hitherto dealt with separate aspects of social behavior. (p. 7).

A second and separate matter of dispute was trade unions' perceptions about the functioning of the economy they act in, and, in particular, about the relation between wages and employment. Dunlop (1944), seeking his inspiration in the comparatively advanced theory of the monopoly firm, assumed that the union perceives a negative relation between wages and employment.

Critique of this assumption has followed two lines. It has been argued either that the suggested relation is too complex to be understood by a trade union; or, that the union's basic outlook is more Keynesian, signifying that the union does not believe in any wage-employment trade-off.[11]

Underlying the present study is the view that it is, at least for some purposes, meaningful to treat trade unions as rational economic agents. It does not seem to be inherently more difficult to analyze trade unions than to analyze other organizations that are treated as economic units, such as families, corporations or nations.[12, 13] Studies of these organizations should benefit as much as should trade union analyses by

> "[e]ven the most primitive clichés of politics sociology and psychology..."(Ross(1948, p. 7)).

Acceptance of the notion that traditional economic analysis is, for some questions, applicable to trade unions, does not imply agreement with any particular characterization of trade union behavior. All that is accepted is that their actions may be viewed as rational attempts to achieve some target(s), given certain perceived constraints. The problem of identifying these targets and constraints still remains to be solved.

The next subsection briefly presents various ways of representing trade unions' internal organization and targets, and some approaches to the union-employer relationships are presented in subsection 3.3.

3.2 *The Trade Union's Internal Organization and Objectives*

The formal literature usually depicts the trade union as a *unit*. A manifestation of this is that the organization can be characterized by some objective function. However, since the union comprises a number of individual members, there is a question of exactly how the objective function is derived.[14] The literature here falls into two broad categories.

Studies in the first category assume a union objective function, without making explicit its relation to individual members' preferences. The union is treated as a *"black box"*, such that only the outcome of the internal processes - the union's objective function - is observable. Frequently, these functions are said to represent the union leadership.

A variety of different trade union objectives have been suggested. One of the most commonly adopted in the earlier literature was maximization of the total wage bill of the union membership. It was suggested by Dunlop (1944), who also listed a number of alternative conceivable objectives. Dunlop (1944), Rosen (1970), and de Menil (1971) all propose that the union might seek to maximize the "rent" from unionization. A large membership is sometimes claimed to be an objective, in particular in studies where the preference function is supposed to represent the preferences of the union leadership. The most general approach is to simply postulate an objective function without specifying exactly where it stems from. The student then includes whatever arguments are conceived of as being appropriate; Fellner (1947), Fellner (1949), Cartter (1959), and Rees (1962) constitute early examples of this.

The disadvantage of this *ad hoc* approach is that it is hard to discriminate between different representations on any *a priori* ground - witness the proliferation of proposed union objective functions. Also, it is difficult to predict the change in a union's behavior that results from a change in its economic environment. However, certain preference or reaction functions are very hard to derive from any reasonable micro theory, but may nevertheless appear most reasonable from an empirical point of view.[15]

The second category of literature on trade union behavior takes the stringent requirements for permissible aggregation of individual preferences more seriously. The union preference function is here derived by explicitly taking into account the organizational procedures within the union.[16]

One step in this direction, away from the black box assumption, is taken in e.g. McDonald and Solow (1981) and Oswald (1982). These authors assume the existence of a representative union member, or, which in this case amounts to the same thing, that the union has utilitarian preferences defined over the members' welfare levels. This approach assumes that all members are identical as far as decisions about union actions are concerned. For instance, it requires that the probability of being unemployed is the same for all members. The possible rationing of jobs is assumed to take place via a random draw.[17] However, the assumed existence of a representative member obviously makes this approach less interesting as a description of the political process within the union. On the other hand, it enables an anlytically simple yet not completely *ad hoc* trade union representation.

Not surprisingly, the notion of a representative member has worried some economists sufficiently to make them seek alternative ways of modelling trade unions, ways where diversity of preferences among members plays a prominent role. Farber (1978) and Grossman

(1983a, b) develop models where workers are ranked according to some criterion; in Farber's (1978) case according to the value of the members' subjective discount parameters, and in the case of Grossman (1983a, b) according to seniority. The demand for labor is partly random, and the union must set the wage before this uncertainty is resolved, implying that workers have different most-preferred wages. Decisions in the union are reached through a voting procedure with direct democracy, and the outcome of the vote can be shown to be the *median's* most-preferred wage. This type of model is analytically not as simple to work with as the representative worker representation, but has better microfoundations than any of the other approaches, and it manages to capture some difference in interest among workers.

3.3 *Trade Union-Employer Relationships*

Given a trade union objective function, the next step is to determine the institutional framework within which the union and the employer interact. Several different models have been used in the literature.

One approach is to assume that the union unilaterally determines a wage, leaving it to firms to hire the quantity of labor that maximizes profits at this wage. The union hence chooses a point along the employers downward sloping demand curve for labor. This is sometimes referred to as the *monopoly union* representation because of the close similarity to the representation of the monopoly firm.

A second model originates in a paper by Leontief (1946), and is developed further by e.g. Oswald and Ulph (1982). Here the union not only unilaterally sets a wage and leaves the quantity decision to the employer, but sets *both* the wage *and* the quantity hired, and leaves it to the employer to accept or not. By offering a sufficiently high profit level, the union can ensure that the wage proposal is accepted.[18]

This type of union is strong enough to force the employer off the derived demand curve for labor; that is to hire a non-profit maximizing number of workers. The union acts as a perfectly discriminating monopolist, extracting all the "consumer surplus". To emphasize the union's strong bargaining position, it could be referred to as the "*coercive*" union.

A third model, employed by Fellner (1949), and later used by e.g. McDonald and Solow (1981), explicitly considers the wage formation process as a two-agent bargaining game. This approach is here referred to as the *"bargaining"* union.

There is one major difference between the monopoly union model, on the one hand, and the coercive union and the bargaining union model, on the other: the monopoly union solution is not efficient, as could be expected considering the price *or* quantity setting monopolistic feature. There are usually combinations of lower wages and higher employment levels that both the union and the employer prefer relative to the monopoly solution. Hence, by moving away from the labor demand curve both parties could gain. In the equilibrium case of the coercive or the bargaining union, however, all such gains are realized - the solution is on the contract curve.

One would expect that the inefficiency of the monopoly union solution would make it a less common solution in situations where the union bargains with one, or a few, employers: the gains from moving to the contract curve should in these cases exceed the additional bargaining costs therefor. The monopoly union, on the other hand, seems *a priori* a better representation of cases where the buyer's side consists of a larger number of employers, coordinated through a federation. If both employment levels and wages were to be negotiated, the federation would face severe problems in determining how to allocate the agreed employment. This would require too much information about individual firms to be practicable.[19]

There is also a major difference between the bargaining union model, on the one hand, and the monopoly and coercive union models, on the other: the solution is cooperative in the former case, but non-cooperative in the latter two. The former approach suffers from the usual weaknesses of this sort of cooperative game. First, it is usually difficult to establish the range within which the negotiated wage will lie (i.e. to find the "threat-points"). Second, there are a variety of possible solution concepts, which are all more or less *ad hoc*. Exactly where in the established range the solution will occur is determined by relative bargaining power, and such a concept can be difficult to meaningfully formalize.[20]

Before presenting the contents of chapters V and VI, it should be pointed out that there are several important areas in the literature on trade unions that have not been dealt with here, such as determinants of the degree of unionization, the extent to which there are union-non-union wage differentials, determinants and effects of labor disputes, determinants and effects of the level at which union-employer negotiations are conducted, the impact of trade unions on income distribution, trade unions and macroeconomic performance, etc. Hence, this very brief survey does not give a correct account of the richness of the literature on trade unions.

3.4 *Contents of Chapters V and VI*

Chapters V and VI are both inspired by experiences of large public sectors and strong trade unions in the smaller European market economies, and both chapters consider the relation between these phenomena. A particular focus is on the wage formation process, and on how wages affect total employment in this type of economy. Just as in the macroeconomic disequilibrium literature originating in Patinkin (1965), Clower (1965), and Barro and Grossman (1971), wages are not necessarily market clearing. But in contrast to this literature, wages are here not fixed exogenously, but are endogenously determined. Moreover, no account is taken here of spill-over effects and multiplier effects.

As already mentioned, there are different approaches to the modelling of trade unions, and their external relationships. The description of the trade union's internal organization differs between the two chapters, as discussed below. But in both cases the external relationships are depicted by the monopoly union model, the reason being that in the type of economy portrayed wages have largely been set through centralized bargaining, with employment decisions left to the employers. Clearly, the assumption underlying the monopoly union model, that the wage is set unilaterally by the trade union, is questionable as a depiction of the real world. But it facilitates the analysis computationally and therefore allows complexities to enter from other sources. It also serves to highlight the union's influence on the economy, by removing all strategic behavior from the employers' side.

The monopoly union model is obviously not the only possible approach for analyses of the issues focussed upon here, and applications of the bargaining union model might provide new insights.[21] However, when evaluating the pros and cons of these models, one should take into account the fact that both chapters V and VI are mainly concerned with the direction of *changes* in wages and employment, rather than their levels; i.e. they seek qualitative, rather than quantitative, answers. It remains to be seen whether these variables move in the same or the opposite direction with other approaches. This is an area for further research.

Another somewhat unsatisfactory assumption for analyses of economies with centralized wage-setting is the assumption that the trade union perceives a downward sloping demand curve for labor. The influence of Keynesian thinking has been strong, not least among trade unions, and it is therefore questionable whether unions have believed in an aggregate wage-employment trade-off. This is probably a problem in particular for analyses of the period up to the mid 70s. The development of models based on other perceptions than the classical demand curve for labor is therefore another important area for future research.[22]

The purpose of chapter V is to study some aspects of the relation between the size of the public sector and total unemployment in an economy with a private and a public sector where all privately employed workers are organized in a union that sets the private sector wage. The analysis is inspired by the present discussion of the actual and the appropriate role of the public sector in today's economies.

The chapter's basic idea is that the divergence in interest between more and less senior workers might be of importance for questions involving the movement of workers from one sector to the other. To capture these differences in interests, a slightly altered version of the median-voter model developed by Grossman (1983a, b) is employed.

Three points are made in the chapter. First, it is shown that in this economy the long run unemployment rate is unambiguously lower, the larger the public sector, for a given income tax rate on workers. Second, it is pointed out that there might exist a ratchet-type effect in the wage setting process, in the sense that the wage increases immediately in response to an increase in the size of the public sector, whereas the wage adjusts sluggishly downwards when the public sector contracts. As a consequence, a once-and-for-all reduction in the size of the public sector might cause "longer" term and possibly even permanent unemployment. However, it is argued that by using appropriate tax policies, the unemployment could be reduced. Finally, it is shown that an employment stabilization policy that is based on public sector interventions into the labor market, not only stabilizes employment, but might also have long run allocational effects, in that it reduces expected long run private employment.

Chapter VI considers implications for wages and employment of an employment stabilization policy that works through direct increases in public sector employment or direct subsidization of private firms. The focus is, in particular, on how the trade union's awareness of this employment policy affects the union's wage decision. The wage level determines the private sector employment, and might therefore in turn trigger off public sector stabilization measures.

Since the issue of main interest is the interrelation between actions undertaken by the government and the union, and not the union's internal political processes, the representative member model is used to describe the union, for analytical convenience. This choice is, however, not crucial; other union representations would give qualitatively the same results.

There is again a private and a public sector, but here the trade union organizes workers in both sectors and it sets a uniform wage for all employed workers. Unemployment is randomly allocated, workers are in all aspects identical, and hence there exists a representative member. This worker has (derived) preferences defined not only over wages and employment, but also over consumption of the public sector's output.

The chapter shows that if the trade union perceives the public sector's stabilization policy, the union will increase the wage until the benefit from the resulting expansion of the public sector is just outweighed by the induced costs in terms of higher taxes. Hence, the policy might lead to a structural transformation of the economy, with a larger public sector, and a correspondingly smaller private sector. Furthermore, since the government is assumed to accommodate only part of the resulting unemployment, the employment stabilization policy tends to increase overall unemployment. Similar effects can also arise if the trade union's and the government's goals for employment differ.

4. FINAL REMARKS

There are some similarities between the type of model employed in this study to portray trade unions and monopolistically competitive firms. All these models describe price or quantity setting organizations, that optimize against downward sloping demand curves. Firms maximize profits and trade unions maximize utility.[23] A firm's profit hinges on the supplied quantity both directly, since it determines the quantity that is sold, and indirectly by influencing the received price and the incurred production cost. Similarly, in many trade union models, the union's utility depends on its "supplied" quantity directly, and indirectly by influencing the received wage and the incurred cost in terms of foregone leisure.[24] These union models are hence very similar to the monopoly firm model. The differences are that the "cost" side is usually simpler in trade union models (in fact, in this study there is no cost at all from foregone leisure), whereas the "revenue" side is substantially more complicated.

It is sometimes claimed that in any reasonable model of a trade union, the size of the union's membership should be a source of utility, since a large membership renders the union leaders prestige. The argument probably has some validity and shall not be disputed. But it deserves to be pointed out that a very similar argument is often raised against the model of the profit maximizing firm. The assumptions underlying this model indeed appear naive, in particular today when ownership and management are most often divided in larger firms. There is a large number of organization theorists who suggest other objectives for firms. For instance, Baumol (1958, 1959) argues that firms strive to maximize sales revenues, subject to certain profit constraints, and Williamson (1964) claims that managers' utility functions increase in e.g. expenditures on staff. In sum, it is bard to see why managers should be less concerned than union leaders with personal status, or why such status should be inherently more tied to the size of the union leader's organization, than to the size of the business leader's organization.

While this argument is at best only a weak defense of the trade union models in chapters V and VI, it might nevertheless be worth mentioning, if it reminds the reader how simplistic the treatment of firms is in traditional economic theory, and that a bad assumption does not become any better by being repeated *ad infinitum*.

Finally, some readers may be disturbed by the fact that models and results are not exposed to the real world e.g. through simulations or empirical testing. However, the analyses in the following five chapters are unfortunately not yet sufficiently refined for these purposes. They are also still far too simple to permit inferences about the real world. Instead, the models should be seen as explorations into *analytical possibilities* rather than into e.g. policy prescriptions. Their main virtue is their technical simplicity. By removing from the analysis as many less important aspects as possible, these models serve to unveil the relationships of particular interest, making the economics rather than the mathematics or econometrics the crux of the matter. The value of this approach critically hinges on the fact that crucial aspects are not assumed away. Whether or not this is done in the present study is, however, left to the reader to judge.

FOOTNOTES

* I am indebted to Lars Calmfors, Harry Flam, Gene Grossman, Assar Lindbeck, Torsten Persson, Peter Svedberg, Lars Svensson and Hans Tson Söderström for many helpful comments and suggestions.

1. Intra-industry trade implies that a country simultaneously exports and imports similar goods, something which does not occur according to traditional trade theory. The estimate of 50% is given by Grubel and Lloyd (1975).

2. See Hood and Young (1979).

3. Most of the material considering the period up to the present century is taken from Schumpeter (1954).

4. See the Preface to the first edition, reprinted in Marshall (1920).

5. Game Theory seems to have undergone a similar cycle; a lot of developments in the 50s, declining attention in the 60s and 70s, and a renewal in interest in the early 80s.

6. It is interesting to note that both the early 30s, and the late 70s, which were periods when the interest in imperfect competition increased, were preceded by a period of volatile economic developments. Maybe this results from the fact that price or wage rigidities are more easily noticed, and probably have stronger impacts, when the economic environment undergoes drastic changes, putting pressure on relative prices?

7. The discussion here is confined to horizontal product differentiation only, and hence not to vertical differentiation where pro-

ducts differ in their quality. For a discussion of this in the context of trade in differentiated products, see e.g. Gabszewicz et al (1981).

8. For a discussion of the pros and cons of the Spence-Dixit-Stiglitz and the Lancaster representations, see e.g. Pettengill (1979), Dixit and Stiglitz (1979), and Lancaster's comment in Krugman (1982).

9. Since I agree with Schumpeter's (1954) view that

> ....[Mill] is so familiar to every educated person that it might seem superfluous to add anything to what can be read in dozens of books... (p. 527),

I shall avoid insulting the reader by refraining from further mention of Mill.

10. See Mitchell (1972) for other references to participants in this debate.

11. Recent criticism of the assumption about a trade union perceived downward sloping demand curve for labor, can be found in e.g. Mitchell (1980, pp. 64-77).

12. One can, of course, argue perfectly consistently (as some would) that traditional economic analysis has little or nothing to contribute to the understanding of these entities, either.

13. Farber (1978), Dertouzos and Pencavel (1981), and Pencavel (1982) attempt to empirically estimate union preference functions.

14. See e.g. Cartter (1959), or Atherton (1973), for discussions of union objectives.

15. The fundamental question whether a model's realism should lie in its assumptions or results is discussed by e.g. Friedman (1953).

16. In this case we have to be careful not to "personalize" the union by giving it preferences. The essence of the approaches within this category is that the trade union is exactly what it says: a *union* of individuals.

17. It is assumed that the number of working hours per employed worker is constant.

18. There is hence an element of cooperation here. But, this feature does not seem strong enough to warrant this model to be denoted "cooperative", as suggested by Oswald and Ulph (1982).

19. Hall and Lillien (1979) show a case where unilateral employment decisions by a firm lead to efficient wage bargains.

20. See e.g. de Menil (1971, ch. 2) for a discussion of various bargaining theories.

21. Ellis and Fender (1982) present an interesting application of the bargaining approach.

22. Gylfason and Lindbeck (1982) is a step in this direction.

23. As mentioned above, one should be careful not to misuse the utility interpretation of the voting union portrayed in chapter V.

24. E.g. Oswald (1982), and Calmfors (1982) work with models of trade unions representing members that derive utility from leisure.

REFERENCES

Aristotle, 341 B.C., "Ethics V", Working Papyrus 17, Department of Philosophy, Lyceum.

Arrow, K. and G. Debreu, 1954, "Existence of an Equilibrium for a Competitive Economy", *Econometrica,* 22, pp. 265-290.

Atherton, W., 1973, *Theory of Union Barganing Goals,* Princeton University Press.

Barro, R. and H. Grossman, 1971, "A General Disequilibrium Model of Income and Employment", *American Economic Review,* 61, pp. 82-93.

Baumol, W., 1958, "On the Theory of Oligopoly", *Economica* 25, pp. 187-198.

Baumol, W., 1959, *Business Behavior, Value and Growth,* Macmillan.

Calmfors, L., 1982, "Employment Policies, Wage Formation and Trade Union Behavior in a Small Open Economy", *Scandinavian Journal of Economics,* No. 2.

Cartter, A., 1959, *Theory of Wages and Employment,* Irwin, Illinois.

Chamberlin, E.H., 1933, *The Theory of Monopolistic Competition,* Harvard University Press, Cambridge, Massachusetts.

Clower, R., 1965, "The Keynesian Counter revolution: A Theoretical Appraisal", in Hahn, F. and F. Brechling (eds), *The Theory of Interest Rates,* Macmillan, London.

Cournot, A., 1960, *Researches into the Mathematical Principles of the Theory of Wealth,* Kelley, New York.

Dertouzos, J. and J. Pencavel, 1981, "Wage and Employment Determination under Trade Unionism: The International Typographical Union", *Journal of Political Economy,* 89, No. 6, pp. 1162-1181.

Dixit, A. and J. Stiglitz, 1977, "Monopolistic Competition and Optimum Product Diversity", *American Economic Review,* 67, No. 3, pp. 297-308.

Dixit, A. and J. Stiglitz, 1979, "Monopolistic Competition and Optimum Product Diversity: Reply", *American Economic Review,* 69, No. 5, pp. 961-63.

Dunlop, J., 1944, *Wage Determination under Trade Unions*, Macmillan, New York.

Eaton, J. and H. Kierzkowski, 1983, "Oligopolistic Competition, Product Variety and International Trade", in Kierzkowski, H. (ed), *Monopolistic Competition and International Trade*, Oxford University Press, (forthcoming).

Ellis, C. and J. Fender, 1982, "Wage Barganing in a Macroeconomic Model with Rationing", Working Paper No. 279, Dept. of Economics, Cornell University, New York.

Farber, H., 1978, "Individual Preferences and Union Wage Determination. The Case of the United Mine Workers", *Journal of Political Economy*, 86, No. 5, pp. 923-942.

Fellner, W., 1947, "Prices and Wages under Bilateral Monopoly", *Quarterly Journal of Economics*, Aug., pp. 503-532,

Fellner, W., 1949, *Competition Among the Few*, A.A. Knopf, New York.

Friedman, M., 1953, "The Methodology of Positive Economics", in *Essays in Positive Economics*, University of Chicago Press, Chicago, pp. 1-43.

Gabszewicz, J., A. Shaked, J. Sutton and J.-F. Thisse, 1981, "International Trade in Differentiated Products", *International Economic Review*, 22, No. 3, pp. 527-534.

Grossman, G., 1983a, "Union Wages, Temporary Layoffs and Seniority", *American Economic Review* 73, No. 3, (forthcoming).

Grossman, G., 1983b, "International Competition and the Unionized Sector", Seminar Paper 251, Institute for International Economic Studies, University of Stockholm.

Grubel, H. and P. Lloyd, 1975, *Intra-Industry Trade: The Theory and Measurement of International Trade in Differentiated Products*, Macmillan, London.

Gylfason, T. and A. Lindbeck, 1982, "The Macroeconomic Consequences of Endogenous Governments and Labor Unions", Seminar Paper No. 232, Institute for International Economic Studies, University of Stockholm.

Hall, R. and D. Lillien, 1979, "Efficient Wage Bargains under Uncertain Supply and Demand", *American Economic Review*, 69, No. 5, pp. 868-879.

Helpman, E., 1983, "Increasing Returns, Imperfect Markets, and Trade Theory", in Jones, R. and P. Kenen (eds) *Handbook of International Trade*, North-Holland, (forthcoming).

Hood, N. and S. Young, 1979, *The Economics of Multinational Enterprise*, Longman, London.

Krugman, P., 1982, "Trade in Differentiated Products and the Political Economy of Trade Liberalization", in Bhagwati, J. (ed), *Import Competition and Response*.

Lancaster, K., 1979, *Variety, Equity, and Efficiency*, Columbia University Press, New York.

Lancaster, K., 1980, "Intra-industry Trade under Perfect Monopolistic Competition", *Journal of International Economics*, 10, pp. 151-175.

Leontief, W., 1946, "The Pure Theory of the Guaranteed Annual Wage Contract", *Journal of Political Economy*, 54, pp. 76-79.

McDonald, I., and R. Solow, 1981, "Wage Bargaining and Employment", *American Economic Review*, 71, No. 5, pp. 869-908.

Marshall, A., 1920, *Principles of Economics*, 8th ed., Macmillan, London.

de Menil, G., 1971, *Bargaining: Monopoly Power Versus Union Power*, M.I.T. Press, Cambridge, Massachusetts.

Mill, J., 1909, *Principles of Political Economy with Some of Their Applications to Social Philosophy*, 7th ed., London.

Mitchell, D., 1972, "Union Wage Policies: The Ross-Dunlop Debate Reopened", *Industrial Relations*, 11, pp. 46-61.

Mitchell, D., 1980, *Unions, Wages, and Inflation*, Brookings Institution, Washington.

Oswald, A., 1982, "The Microeconomic Theory of the Trade Union", *Economic Journal*, 92, pp. 576-595.

Oswald, A. and D. Ulph, 1982, "Unemployment and the Pure Theory of the Trade Union", Oxford University and University College, London, mimeo.

Patinkin, D., 1965, *Money, Interest and Prices*, 2nd. ed., Harper and Row, New York.

Pencavel, J., 1982, "The Trade-Off Between Wages and Employment in Trade Union Objectives", Working Paper No. 870, National Bureau of Economic Research, Cambridge, Mass.

Pettengill, J., 1979, "Monopolistic Competition and Optimum Product Diversity: Comment", *American Economic Review*, 69, No. 5, pp. 957-60.

Rees, A., 1962, *The Economics of Trade Unions*, University of Chicago Press.

Robinson, J., 1933, *The Economics of Imperfect Competition*, Macmillan, London.

Rosen, S., 1970, "Unionism and the Occupational Wage Structure in the United States", *International Economic Review*, 11, pp. 269-286.

Ross, A., 1948, *Trade Union Wage Policy*, California University Press, Berkeley.

Salop, S., 1976, "Information and Monopolistic Competition", *American Economic Review*, 66, No. 2, pp. 240-245.

Schumpeter, J., 1954, *History of Economic Analysis*, Allen and Unwin, London.

Spence, A.M., 1976, "Product Selection, Fixed Costs, and Monopolistic Competition", *Review of Economic Studies*, 43, No. 2, pp. 217-35.

Sraffa, P., 1926, "The Laws of Returns Under Competitive Conditions", *Economic Journal*, 36, pp. 535-550.

Takayama, A., 1974, *Mathematical Economics*, Dryden Press, Illinois.

Viner, J., 1950, *The Customs Union Issue*, Carnegie Endowment for International Press, New York.

Walras, L., 1954, *Elements of Pure Economics: Or the Theory of Social Wealth*, Irwin, Homewood.

Williamson, O., 1964, Economics of Discretionary Behavior: Managerial Objectives in a Theory of the Firm, Kershew.

CHAPTER II

SOME IMPLICATIONS OF NON-HOMOTHETICITY IN PRODUCTION IN A TWO-SECTOR GENERAL EQUILIBRIUM MODEL WITH MONOPOLISTIC COMPETITION

CHAPTER II SOME IMPLICATIONS OF NON-HOMOTHETICITY IN PRODUCTION IN A TWO-SECTOR GENERAL EQUILIBRIUM MODEL WITH MONOPOLISTIC COMPETITION*

1. INTRODUCTION

Although it has been well known for decades that a substantial proportion of international trade can be characterized as intra-industry trade, this phenomenon has until recently defied formal general equilibrium analysis. However, recently a number of studies have been presented which develop formal models which are able to generate intra-industry trade.[1]

With some exceptions these studies typically develop two-sector, two-factor, two-country models. One sector is of the traditional constant returns to scale type. The novelty is the introduction of increasing returns to scale, product differentiation, and Chamberlinian monopolistic competition in the other sector. Although these models sometimes differ in their representation of the demand side, they share the same basis for intra-industry trade: economies of scale at the firm level. With some demand for all specifications in both countries, and no variant simultaneously produced by two firms, there will be intra-industry trade, although each individual variant is traded in one direction only.

In these models inter-industry trade is as usual explained by comparative advantage based on relative factor endowments. When the economies of scale approach constant returns, and the elasticity of substitution between the different variants goes to infinity, these models reduce to the traditional two-sector factor proportions model, and can hence be seen as *generalization of the standard Heckscher-Ohlin model*.

The purpose of this chapter is to study how non-homotheticity on the production side influences some properties of this two-sector, two-factor economy.[2,3] Non-homotheticity can, for instance, be due to a (one product) firm simultaneously engaging in different activities, if these activities employ factors in different proportions. Since products are usually differentiated through, for example, marketing or R & D efforts, and since these activities are likely to differ in their factor content from the firm's other activities, it seems that non-homotheticity should be a common feature of differentiated goods industries, and therefore deserves our attention in the context of intra-industry trade.

We are interested in how non-homotheticity in the technology of the differentiated goods industry affects relations between factor endowments, prices and output quantities. As a standard of comparison we will use the comparative statics equations derived by Jones (1965), as these are by now the standard way of expressing the properties of the traditional two-sector model. We will also compare some of our results with some in Jones (1968). In this paper Jones studies some basic trade issues in a two-sector, two-factor economy, where there are economies or diseconomies of scale that are external to the individual, perfectly competing firms, and where non-homotheticity is also permitted. Of special interest to this chapter is Jones' (1968) results concerning the Rybczynski theorem.

As mentioned above, a couple of different representations of preferences have so far been used in the literature. One group, represented by Lancaster (1980) and Helpman (1981), for example, employs a spatial approach where the commodity space is representable by some geometric device, such as a line or a circle, along which the different variants are located. Another approach, used by e.g. Krugman (1979, 1980) and Dixit and Norman (1980) assumes that the representative consumer's preferences over the different variants can be represented by a parametricized utility function of a CES type. A main difference between the two approaches is that with the former an increase in the number of variants will make the product space more crowded, and thereby increase the elasticity of demand, whereas this will not necessarily affect the demand elasticity with the second approach. We have chosen to work with the second type of representation, generalizing the model employed by Dixit and Norman (1980) to include non-homotheticity in the differentiated goods industry, mainly because of its computational convenience.

We assume that the elasticity of substitution in demand between the two sectors is unity, although this means that our model in this respect is less general than the traditional two-sector model.[4] We believe that this is justified, since our main interest is in how the technology of the differentiated goods industry influences the economy's equilibrium. But we should also be aware of the fact that this simplifying assumption has a significant impact on the results. This is due to the fact that in its absence the share of total income that the representative consumer devotes to the differentiated goods industry, depends not only on the relative prive between the two sectors (as in the traditional model), but also on the number of variants in existence. The implication of this is that even though perceived demand elasticities are constant, the absolue size of the economy, as measured by the scaling of the factor endowments, will influence, for instance, the relative factor reward. We therefore let the demand elasticity of substitution between sectors be unity to avoid additional complications entering from the demand side.

Finally, we disregard problems concerning indivisibilities in firm size (integers), and concerning the existence and uniqueness of the equilibrium under study.

The chapter is structured as follows. Section 2 contains a brief description of the model, and a presentation of the general equilibrium system. In section 3 we define our measure of non-homotheticity, and we examine how the relative factor reward affects the representative firm's output volume, factor proportions, and price, in the presence of non-homotheticity. This hopefully facilitates the understanding of the results in section 4, in which we use the general equilibrium system to determine how factor endowments affect relative factor rewards, the number of firms in the differentiated goods industry, and relative industry output volumes. We identify the factors that determine whether, for example, a Rybczynski-type proposition (with variable factor prices) is valid also here, and we study how non-homotheticity affects the economy's sensitivity to factor endowment changes. The final section (section 5) considers the possibility of predicting inter-industry trade patterns from observations of pre-trade equilibrium levels of various endogenous variables.

2. THE MODEL

Consider an economy in which goods are produced in two industries, denoted by X and Y. The X-industry is producing a group of differentiated goods, whereas the commodity manufactured in the other, 'outside', industry is homogeneous. Following Dixit and Stiglitz (1977) we let the representative consumer's preferences be represented by the utility function

$$u = \left(\sum_{i=1}^{n} x_i^{\beta} \right)^{\alpha/\beta} Y^{1-\alpha}; \qquad 0 < \alpha < 1; \qquad 0 < \beta < 1. \qquad (1)$$

where n is the number of variants consumed.

The advantage of this representation is that if the firms behave as Cournot or Bertrand-type oligopolists, and if the number of firms is large, the elasticity of demand for any single variant of the differentiated goods is parametrically given as[5,6]:

$$- \frac{1}{\partial p(\cdot)/\partial x} \frac{p(\cdot)}{x} = \frac{1}{1-\beta} > 1. \qquad (2)$$

We assume that the differentiated goods industry consists of n profit maximizing firms, each producing a similar but, from the consumer's point of view, not identical product. The number of producers is assumed to be sufficiently large so as to warrant the assumption that firms make Cournot or Bernard-type conjectures.

Although there are no formal obstacles to exit or entry, there are economies of scale which limit the equilibrium number of firms. The economies of scale are assumed to be decreasing in the output volume and specific to each variant. Each firm manufactures only one of the variants and is the sole producer of that particular specification. The market structure is Chamberlinian monopolistic competition with vanishing profits in equilibrium.

Producers transform capital and labor into final products with the same technology, representable by a twice differentiable, increasing and strictly quasi-concave production function. Associated with this production function is the cost function $\phi(w, r, x)$, where w and r denote the factor rewards for labor and capital, respectively, and x is the firm's output volume. The assumptions made above about returns to scale imply that for any firm marginal cost has to be less than average cost, and that the elasticity of the ratio between these with respect to output has to be positive, or

$$\theta \equiv \frac{\partial\phi}{\partial x}\frac{x}{\phi} \equiv \frac{x\phi_x}{\phi} < 1, \tag{3a}$$

and

$$\eta \equiv \frac{\partial\theta}{\partial x}\frac{x}{\theta} = 1 - \theta + \tau > 0, \tag{3b}$$

where

$$\tau \equiv \frac{\partial\phi_x}{\partial x}\frac{x}{\phi_x} \equiv \frac{x\phi_{xx}}{\phi_x} .$$

$\theta = MC/AC$ is the partial elasticity of scale with respect to output, and it is used here as a measure of the economies of scale.[7] It takes on values less than unity when there are increasing returns to scale, and unity when there are constant returns. η is the partial elasticity of this measure with respect to output.

Our equilibrium system is the same as in Helpman (1981), with the difference that our assumptions about the demand side imply a constant ratio between average and marginal revenue. We therefore refer to this paper for details, and here simply give the equations. The first-order condition for profit maximization is that marginal revenue equals marginal cost[8]:

$$p\beta = \phi_x(w,r,x). \tag{4}$$

The condition for industry equilibrium is that each firm is breaking even:

$$px = \phi(w,r,x), \tag{5}$$

and hence by (4) and (5):

$$\theta(w,r,x) = \beta. \tag{6}$$

The outside industry Y is a traditional constant returns to scale industry. The price of its good is chosen as the numeraire, and therefore equilibrium requires that

$$1 = b(w,r), \tag{7}$$

where $b(w,r)$ is the unit cost function. Factors of production are inelastically supplied in amounts L and K, and the full employment conditions for a symmetric equilibrium are:

$$n\phi_w(w,r,x) + Yb_w(w,r) = L, \tag{8}$$

and

$$n\phi_r(w,r,x) + Yb_r(w,r) = K. \tag{9}$$

Equilibrium in the goods markets is given by:

$$\frac{n\ p\ x}{Y} = \frac{\alpha}{1 - \alpha}\ . \tag{10}$$

Eqs. (4)-(10) constitute six independent equations that suffice to determine the three relative prices p, w and r; the two quantities x and Y; and the number of variants produced in the differentiated goods industry n.

3. EQUILIBRIUM IN THE X-SECTOR

In this section we will, before studying the full general equilibrium system, look at three aspects of the differentiated goods industry, namely how the relative factor price affects the representative firm's output volume, factor proportion, and price. These relations are either immaterial to, or different in, the traditional two-sector production model, where the production function exhibits homotheticity and constant returns to scale. Therefore, in order to facilitate interpretations of the results in the following section, some preliminaries are required.

The technology of firms in the differentiated goods industry differs in two respects from what is assumed in the traditional Heckscher-Ohlin model. First, we have increasing, and variable, returns to scale here. To capture these phenomena we defined the two variables θ and η. Second, we allow here also for non-homotheticity, and to capture this we now introduce the variables:

$$\mu_L = \frac{x\phi_{wx}(w,r,x)}{\phi_w(w,r,x)} \quad \text{and} \quad \mu_K = \frac{x\phi_{rx}(w,r,x)}{\phi_r(w,r,x)} .$$

μ_j is thus the partial elasticity of demand for factor j with respect to output, which is assumed to be positive. We have that[9]:

$$\theta = \theta_{LX}\,\mu_L + \theta_{KX}\,\mu_K , \tag{11}$$

where θ_{jx} is the traditional distributive share of factor j. When the technology is homothetic, $\theta = \mu_L = \mu_K$. We can therefore measure the degree of non-homotheticity for a factor j by the amount its partial elasticity of demand w.r.t. output differs from the scale elasticity θ, i.e. by $(\mu_j - \theta)$. In our two factor case we have that:

$$\theta_{LX}(\mu_L - \theta) = -\,\theta_{KX}(\mu_K - \theta), \tag{12}$$

and it is hence immaterial whether we use $(\mu_L - \theta)$ or $(\mu_K - \theta)$ as our measure of non-homotheticity.[10]

The most immediate implication of non-homotheticity is of course that a firm's factor proportions change as output changes, assuming a constant relative factor price. The relation between our measure of non-homotheticity and the change in distributive shares is indeed very close:

$$\mu_j - \theta = \frac{\partial\theta_{jX}}{\partial x} \frac{x}{\theta_{jX}}, \tag{13}$$

that is, the degree of non-homotheticity of factor j is equal to the partial elasticity of that factor's distributive share w.r.t. the output volume.

Let the first comparative statics experiment be the effect on the individual firm's output volume of a change in the wage-rental ratio. To get a link between changes in relative prices and in factor rewards we differentiate (6):

$$\hat{x} = -\frac{\theta_{LX}(\mu_L - \theta)}{\theta\eta} (\hat{w} - \hat{r}) \tag{14}$$

Examining (14) we notice first that in the homothetic case the equilibrium output volume of the individual firm is independent of factor prices. This stems from the fact that in this case our measure of economies of scale is just a function of the output volume, and therefore the equilibrium output volume is dependent of factor prices, and is hence fixed. The equilibrating mechanism which will keep profits at zero is exit and entry, as this will affect profits through changes in the price.

In the case of non-homotheticity, changes in factor prices must affect the representative firm's output volume. From (14) we have that *if there is a local bias towards a relative increase in the use of the factor that has become relatively cheaper (more expensive) the equilibrium output volume of the representative firm will increase (decrease).*

The impact of a factor price change on the factor proportions of a firm in the differentiated goods industry can be seen clearer if the effect is decomposed into two parts, as follows. Assume that the firm and the industry are in equilibrium and that the former is producing at E_o (see Fig. 1) with the total factor intensity F_o. Now let the wage-rental ratio increase; this will imply a new expansion path for the firm. The increase will lead to a traditional *substitution* effect at the old production volume, which will increase the capital-labor ratio to F_1.[11] But with non-homotheticity the optimal production volume must also change and so must the factor proportions once more.

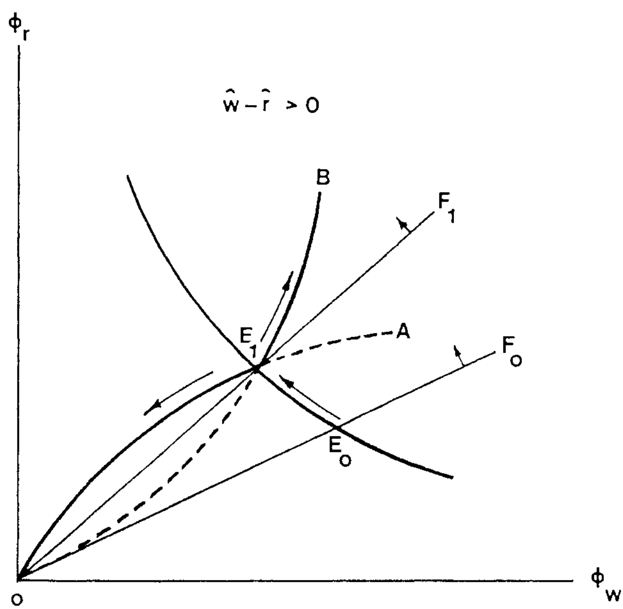

Figure 1

The equilibrium output volume might change in either direction. But for it to decrease there must be a labor-using bias, according to (14), and hence the output effect will move the firm towards the origin along the expansion path OA. Conversely, for the output volume to increase, the bias must be towards capital. The movement will now be outwards along OB. Hence, the output effect must move the firm to a point somewhere on the line that combines the continuous parts of lines OA and OB. Therefore, *regardless of the direction of the non-homothetic bias, the output effect will strengthen the substitution effect.* Formally, the change in factor proportions can be shown as the relative change in the capital-labor ratio:[12]

$$\frac{d(\phi_r/\phi_w)}{\phi_r/\phi_w} = \sigma_X(\hat{w} - \hat{r}) - \frac{(\mu_L - \theta)}{\theta_{KX}}\hat{x}, \tag{15a}$$

where $\sigma_X \equiv \frac{\phi\,\phi_{wr}}{\phi_w\,\phi_r}$ is the elasticity of substitution between labor and capital in the X-sector; or by (14)

$$\frac{d(\phi_r/\phi_w)}{\phi_r/\phi_w} = \left(\sigma_X + \frac{\theta_{LX}}{\theta_{KX}}\,\frac{(\mu_L - \theta)^2}{\theta\eta}\right)(\hat{w} - \hat{r}). \tag{15b}$$

From (15a) and (15b) we have that the relative change in factor proportions that is due to a change in the relative factor price, is larger the more pronounced the degree of non-homotheticity, since the more significant will be the impact of the change in the output volume, *ceteris paribus*.

Let us end this section by deriving the effect of a change in the relative factor price on the price of a differentiated good. We differentiate (5) and use (14) to get

$$\hat{p} = \left(|\theta| + (1 - \theta) \frac{\theta_{LX}(\mu_L - \theta)}{\theta\eta} \right)(\hat{w} - \hat{r}), \tag{16}$$

where $|\theta| \equiv \theta_{LX} - \theta_{LY}$ (to be distinguished from the variable θ).

In the homothetic case a rise in the wage-rental ratio must unambiguously result in a lower (higher) relative commodity price in the X-sector, if it on *average* is capital (labor) intnesive. This stems from that the production volume per firm is constant (cf. our discussion above), while the intensively used factor has become cheaper relative to the other sector. The price must also unambiguously fall when the wage-rental ratio increases if there is a non-homothetic bias towards capital. Finally the price might move in either direction when there is a local bias towards use of the factor that is on average extensively used, relative to the other sector. The second term in (16) will then have the opposite sign to $|\theta|$ and might under certain circumstances be larger in absolute value.

4. RELATIONSHIPS BETWEEN SOME ENDOGENOUS VARIABLES AND FACTOR ENDOWMENTS

We will in this section examine how the relative factor endowment influences the economy's autarky equilibrium when there is non-homotheticity in the production technology of the differentiated goods. When possible, we will compare our results, with the corresponding results in the traditional two-sector economy, as derived by Jones (1965). To do this we must take the full general equilibrium system into account.

So far we have made first-order approximations of the profit maximizing and zero-profit conditions for the two industries, so there remain the equilibrium conditions for the factor and goods markets. The full system can, when differentiated, compactly be stated as:

$$\begin{bmatrix} 0 & 0 & 0 & \theta_{LY} & \theta_{KY} & 0 \\ -1 & 0 & 0 & \theta_{LX} & \theta_{KX} & -(1-\theta) \\ 0 & \lambda_{LY} & \lambda_{LX} & -\delta_L & \delta_L & \lambda_{LX}\mu_L \\ 0 & \lambda_{KY} & \lambda_{KX} & \delta_K & -\delta_K & \lambda_{KX}\mu_K \\ 1 & -1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & \omega_L & \omega_K & 1 \end{bmatrix} \begin{bmatrix} \hat{p} \\ \hat{Y} \\ \hat{n} \\ \hat{w} \\ \hat{r} \\ \hat{x} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \hat{L} \\ \hat{K} \\ 0 \\ 0 \end{bmatrix} \qquad (17)$$

where $\delta_L \equiv \lambda_{LX}\, \theta_{KX}\, \sigma_X + \lambda_{LY}\, \theta_{KY}\, \sigma_Y$,

$$\omega_L \equiv \frac{\theta_{LX}\, (\mu_L - \theta)}{\theta\eta},$$

$$\lambda_{LX} \equiv \frac{n\phi_w}{L},$$

and symmetrically for K. σ_Y is the elasticity of substituion in production in the Y-sector, and λ_{jm} is, as usual, industry m's share of the total usage of factor j.

The determinant of the matrix on the LHS is

$$\Delta \equiv \delta_L + \delta_K + |\lambda||\theta| + \psi \frac{\theta_{LX}}{\theta_{KX}} \frac{(\mu_L - \theta)^2}{\theta\eta},$$

where $|\lambda| \equiv \lambda_{LX} - \lambda_{KX}$,

and $\psi \equiv \lambda_{LX}\, \theta_{KX} + \lambda_{KX}\, \theta_{LX}$.

The determinant is positive regardless of average or marginal factor intensities, and in the homothetic case it simplifies to $(\delta_L + \delta_K + |\lambda||\theta|)$[13].

Let us first consider the relationship between the relative factor reward and the relative factor endowment. It is shown by Helpman (1981) that when preferences over the two industries' goods are representable by a Cobb-Douglas function, and when the technology of the differentiated goods is homothetic, a capital-rich country has a higher wage-rental ratio in an autarky equilibrium than has a more labor abundant country. But is this also true when we allow for non-homotheticity in the X-sector technology? Solving for the relative change in the wage-rental ratio we get

$$\hat{w} - \hat{r} = - \Delta^{-1}(\hat{L} - \hat{K}). \tag{18}$$

In the homothetic case this is exactly the same equation as the corresponding one derived by Jones (1965) for a unitary elasticity of substitution in demand.[14] It is clear, as Δ is always positive, that *an increase (decrease) in the relative endowment of a factor must always decrease (increase) its relative rental, regardless of whether there is a capital or labor using bias in the X-sector technology*. Furthermore, from the definition of Δ we have that *the wage-rental ratio becomes less sensitive to changes in factor endowments the more non-homothetic the technology in the X-sector is, ceteris paribus, regardless of the direction of this bias.* Suppose there is an increase in the supply of, say, capital. The additional capital will enter an economy in a state of equilibrium, and must therefore be supplied at a price lower than the prevailing market price, to become employed. The question is then: How much must the wage-rental ratio increase in order to establish a new full employment equilibrium? This depends as usual on the difference in average factor proportions, and on the elasticity of substitution in production in both sectors. But in addition it will also depend on the output effect that was shown in the previous section to unambiguously strengthen the traditional substitution effect (see Fig. 1). The more non-homothetic the technology the larger will this output be, and hence the less the difference in factor prices that is required to change the representative firm's, and thereby the whole industry's, factor proportions.

The relation between the representative firm's output volume and the relative factor endowment is now immediate from (14) and (18):

$$\hat{x} = \Delta^{-1} \frac{\theta_{LX}(\sigma_L - \sigma)}{\theta\eta} (\hat{L} - \hat{K}) \qquad (19)$$

A more capital (labor) abundant economy will produce more of each variant of the differentiated goods, if the technology is capital (labor) biased at the margin. This is true regardless of whether the sector is more or less capital intensive in marginal or total cost, as compared to the outside industry. The only thing that matters is the difference between the marginal and the total (or average) factor intensity within the firm.

The number of firms is, of course, not only dependent upon the relative factor endowment, but also on the size of the economy. This is easily obtained if we differentiate $p = \alpha I/nx$, which is the symmetric version of the demand function for the representative differentiated good, and use (16), (18), and (19):

$$\hat{n} = \Delta^{-1}\left((1-\alpha)\,[illegible] - \frac{[illegible]}{\sigma}\right)(\hat{L}-\hat{K}) + \gamma_L\hat{L} + \gamma_K\hat{K}, \quad (20)$$

where γ_j is the share of factor j in national income.[15] If there is a proportional change in the endowment of both factors, the number of firms will change in the same proportion and direction, regardless of whether or not the technology is homothetic. This is due both to the unitary elasticity of substitution between the two industries, and to the constant elasticity of demand for any single variant of the differentiated goods.[16]

It is hard, even in the homothetic case, to make any definite predictions about how non-proportional changes in factor endowments change the number of firms. To simplify, we consider changes such that national income is constant at initial prices, so that the sum of the last two terms of (20) is zero. With a homothetic technology the relative total factor intensity then determines whether the number of firms is positively or negatively related to the economy's capital-labor ratio. If the differentiated goods industry is, say, capital intensive, and there is an increase in the relative endowment of capital, the number of firms will increase.

The whole adjustment comes through a change in the *number* of firms.

But, we know from (19) that, in the presence of non-homotheticity, also the output per firm is affected by factor endowment changes. This may increase or decrease the necessary adjustment in the number of firms, relative to the homothetic case. For it to reduce the required adjustment the bias must be towards the factor that on average is relatively extensively used. In this case part, or maybe all, of the adjustment comes through a change in the representative firm's output volume. It is even possible that the number of firms in the differentiated goods industry declines, as a result of an increase in the relative endowment of the factor the industry uses relatively intensively on average.

We now turn to the relation between the relative industry output volume and the relative factor endowment. If we subtract one of the factor market equilibrium equations from the other (as given by (17)), we get that[17]

$$\hat{X} - \hat{Y} = \frac{1}{|\lambda|} (\hat{L} - \hat{K}) + \frac{\delta_L + \delta_K}{|\lambda|} (\hat{w} - \hat{r}) - \frac{\nu}{|\lambda|} \hat{x} , \qquad (21)$$

where $X \equiv nx$,

and $\nu \equiv \lambda_{LX} \mu_L - \lambda_{KX} \mu_K - |\lambda|$.

This expression is identical to the corresponding equation in the Heckscher-Ohlin model, except that there is here an additional $\hat{x}$-term.[18] Rewriting and decomposing ν as

$$\nu = \frac{\psi}{\theta_{KX}} (\mu_L - \theta) - |\lambda|(1 - \theta), \qquad (22)$$

we see that the impact of this term depends, among other things on the extent to which there are economies of scale and non-homotheticity in the X-sector's technology.

Helpman (1981) shows that a version of the Rybczynski theorem holds in this type of economy, when prices are constant. This is also easily seen from (21), as in this case the second and third term on the RHS vanishes. Furthermore, just as in the traditional two-factor model, there will be a magnification effect associated with changes in factor endowments.[19]

Jones (1968) considers an incompletely specialized, two-factor, two-sector economy, where there are variable returns to scale in both sectors. These economies or diseconomies of scale are external to firms, and depend on industry output volumes, so there is still perfect competition. Jones (1968) also allows for non-homotheticity, but in contrast to here it is the *industry* expansion path that exhibits this bias. This implies that the relative marginal factor intensity crucially determines whether the Rybczynski proposition is valid or not, even though factor prices are confined to be constant.[20] Now, in our model, constant factor prices imply that the representative firm's output volume remains fixed irrespective of changes in factor endowments. The whole adjustment in the differentiated goods industry hence has to come through exit or entry. With economies of scale *internal* to the firms we are therefore, in a sense, back in a traditional two-sector economy with constant returns to scale in the number of firms in the X-industry, and output in the outside industry. It is in this case immaterial whether the technology is homothetic or not - only *average* intensities matter. The economy investigated here therefore seems to have, at least from this point of view, properties more similar to the traditional constant returns model than to the external economies or disecononomes to scale model studied by Jones (1968).

How much can be said about the relative output of the two industries when prices are allowed to adjust to the change in factor endowments? We solve for the relative change in the ratio between the two volumes:

$$\hat{X} - \hat{Y} = \Lambda^{-1}\left[|\theta| + (1 - \theta) \frac{\theta_{LX}(\theta_L - \theta)}{\theta_K} \right] (\hat{L} - \hat{K}). \qquad (23)$$

Hence, in the case of homotheticity, the relatively more capital (labor) abundnat economy will produce relatively more in the capital (labor) intensive industry even though prices are variable. In this case eq. (23) reduces to what Jones (1965) derives for the traditional two-sector model, with a unitary elasticity of substitution between sectors.[21]

When we allow for non-homotheticity, life, or at least our model, becomes more complicated. But *we are able to unambiguously sign the expression if the non-homothetic bias is towards the factor that is intensively used on average*. This condition can be interpreted with the help of an Edgeworth box-diagram. Fig. 2 illustrates the case where the differentiated goods are relatively capital intensive ($|\theta| < 0$), and have a bias towards even heavier reliance on capital as output expands, and where in equilibrium there are two firms, A and B, producing in this sector.

Our condition requires that the expansion path of a firm in the differentiated goods industry, at the equilibrium point and at constant factor prices is curved away from the diagonal of the box.

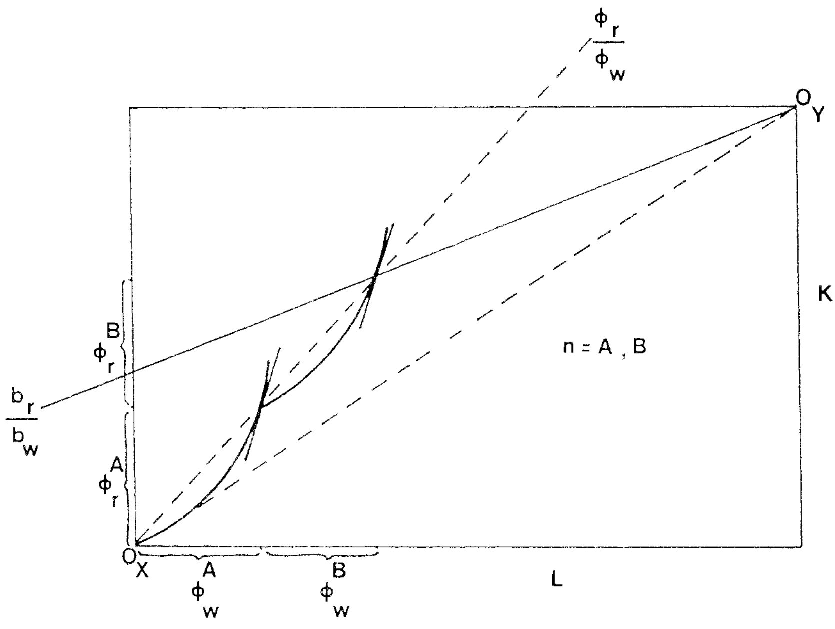

Figure 2

More ambiguous is the case when there is a local bias, as output expands, towards the use of the factor that, relative to the other sector, is extensively used on average.

The second term within the square brackets in (23) will then have the opposite sign to $|\theta|$, and might under certain circumstances be larger in absolute value. Consider the case where the economy's capital-labor ratio increases causing the wage-rental ratio to increase, and where the differentiated goods industry is capital intensive relative to the other sector ($\theta_{LX} > \theta_{LY}$). Assume now that a firm in the X-sector at *the margin* is more *labor* intensive than the Y-sector. The change in the factor price ratio, while consistent with break-even in the outside industry, will result in an increase in the marginal cost at the old production volume, for a typical firm in the X-sector. To restore its equilibrium the firm must, before any entry has occurred, reduce the production volume, which will (regardless of the sign of τ) increase the price. As the firm on *average* is more capital intensive than the Y-sector, the change in the wage-rental ratio must also create positive profits. But if the difference in average factor intensities is small, profits will not increase much, and the entry of new firms will not be very significant. There are thus conceivable cases where the increase in the number of firms is outweighed by the decrease in the production volume of each, so that nx actually falls relative to Y. That is to say, *if the difference between marginal intensities is large, and of the opposite direction compared to average intensities, a relative increase in the supply of a factor might lead to a relative decline in the output volume of the industry that on average employs the factor relatively intensively. This is, ceteris paribus, more likely the more economies of scale there are, and the less sharply they decline in the output volume.*

To end this section, let us just note that if we start in a situation where the technology is homothetic, and introduce a marginal amount of, say, capital biased non-homotheticity, the term multiplying (L - K) in (23) will in absolute terms increase, if the X-sector is capital intensive on average. Hence, *a marginal increase in non-homotheticity will, ceteris paribus, increase (decrease) the sensitivity of the relative industry output volume to changes in the relative factor endowment, if the introduced bias is towards the factor that is on average intensively (extensively) used in the X-sector.* Furthermore, the magnification effect that was unambiguously verified to exist in the homothetic case will be weakened or even cease to exist when the non-homotheticity has the above-mentioned bias. But, on the other hand, if the bias is such that an expansion of the firm's output volume tends to close the gap in average factor intensities, the magnification effect might be strengthened.

5. AUTARKY VALUES AS PREDICTORS OF INTER-INDUSTRY TRADE PATTERNS

In this final section we consider the possibility of predicting trade patterns in a world consisting of two trading economies of the above examined type. We will not present the complete international equilibrium system explicitly; it can be found in different versions in Dixit and Norman (1980) and Helpman (1981), for example. Helpman (1981) shows that a version of the Heckscher-Ohlin theorem for trade patterns is valid, despite the presence of non-homotheticity, if the output effect does not reverse the relative total factor intensity of the two industries. With homothetic and identical preferences in both countries, factor proportions in consumption will be the same, and the relatively capital abundant economy must be a net exporter of the capital intensive good. Relative factor endowments could hence be used as predictors of inter-industry trade patterns, but what about autarky commodity prices, factor rewards, etc?

In the following we assume the above-mentioned condition for factor intensity non-reversals to be fulfilled. We know from (18) that the wage-rental ratio is monotonously related to the capital-labor ratio, and hence the wage-rental ratio is a valid predictor of the inter-industry trade pattern. This is also the case in Helpman (1981), who argues that this is only true when preferences have a unitary elasticity of substitution between the two sectors. Generally, budget shares depend on the number of variants in existence, and hence on the size of the economy.[22] The whole equilibrium, including the relative factor reward, therefore generally depends not only on factor proportions, but also on the size of the economy. Hence, in the absence of a Cobb-Douglas utility function over the two industries, only factor endowments could be used to predict trade patterns.

The relation between the individual firm's autarky output and the relative factor endowment is determined by the type of non-homothetic bias that is present. Therefore, if one knows the direction of the bias, the autarky output volume can be used to predict inter-sectoral trade patterns, if either the countries have marginally different factor proportions or the bias is known to never change direction in the relevant range of x. But if this information is unavailable no prediction can be based on the representative firm's output volume.

A lack of this information also invalidates autarky values of relative industry output volumes, and relative commodity prices as predictors,as is clear from (23) and (16). But, if the direction of the bias is known, these two variables can be used, if the bias is towards the on average relatively intensively used factor.

Finally, if production functions are homothetic, all the above-mentioned variables can be used as predictors of trade patterms, with the obvious exception of the number of firms.

FOOTNOTES

* I am particularly indebted to Elhanan Helpman, and an anonymous referee, for helpful specific suggestions, and to Avinash Dixit, Harry Flam, Carl Hamilton, Torsten Persson, Assaf Razin and Lars Svensson, for valuable comments on earlier drafts. The chapter was presented in earlier versions at the University of Warwick, and at the Institute for International Economic Studies, University of Stockholm, and participants at both seminars provided useful comments. I am grateful to the Jacob Wallenberg Research Foundation for financial support.

1. E.g. Dixit and Norman (1980), Ethier (1979), Helpman (1981), Krugman (1979, 1980), and Lancaster (1980). A slightly different approach, using sequential entry, is Eaton and Kierzkowski (1983). The literature is reviewed in Helpman (1983).

2. Any function that can be expressed as a monotonic transformation of a homogeneous function is denoted "homothetic". Intuitively, we will call a function "locally non-homothetic" if, in input space, the slope of its expansion path locally differs from that of a straight line passing through the origin and the production point.

3. Not all the papers referred to in footnote 1 allow for non-homotheticity, but Helpman (1981) is an example of a trade

model in which e.g. factor price equalization occurs, and in which versions of the Rybczynski and Heckscher-Ohlin theorems are shown to be valid, despite the presence of non-homotheticity.

4. A unitary elasticity of substitution is assumed in e.g. Dixit-Norman (1980), Krugman (1980), and parts of Helpman (1981), for example.

5. For details, see Dixit and Stiglitz (1977), or Dixit and Norman (1980).

6. The firm index is omitted as we are only interested in symmetric equilibria. Note also that in the notation there is for convenience no difference made between consumed and produced amount, as we are concerned only with autarky equilibria.

7. Another frequently used measure is the degree of homogeneity of the production function, i.e., the degree of economies of scale are given by the ζ that satisfies:

$$\lambda^{\zeta}F(K, L) = F(\lambda K, \lambda L).$$

It is shown by Hanoch (1975) that both these measures give the same elasticity of scale, but that with non-homotheticity *the change* in (or the elasticity of) the elasticity of scale (i.e., our η) will be different. Hanoch argues that the concept employed here is the more relevant one for microeconomic analysis, as it assumes cost minimization at all levels of production.

8. The second-order condition is that $\tau(x) > -\frac{1}{\varepsilon}$, which is fulfilled due to the assumption that $\eta > 0$.

9. We use that the marginal cost function is linearly homogenous in the factor prices.

10. We will consequently use the terminology that there is a "bias" towards the use of one of the factors when, as output expands, that factor is increasingly intensive employed relative to the other factor.

11. For simplicity we have in Fig. 1 assumed that the substitution effect brings the firm to the same point E_1 regardless of the direction of the bias in the technology.

12. To arrive at this expression we use homogeneity properties of the cost and factor demand function.

13. This is exactly Jones' (1965) aggregate elasticity of substitution for the case where there is a unitary elasticity of substitution in demand between the outside industry and the utility index over the differentiated goods.

14. Jones (1965), equation (15).

15. To arrive at (20) we use also that $\theta_{LX}\alpha + \theta_{LY}(1 - \alpha) = \gamma_L$.

16. In chapter III we examine a model with the same demand representation as here, but where the number of firms isn't stricly proportional to the size of the economy. Just as in e.g. Helpman (1981) the elasticity of demand facing any single seller increases with the number of firms active on the market, but in contrast to the latter paper, this results from non-Cournot-Nash conjectures, and from avoiding the for smaller markets less satisfactory approximation that $1/n = 0$.

17. The use of X instead of nx can be thought of as only a matter of notational convenience. But in the interpretations of the results below we also use X to denote "the output of the differentiated goods industry". This is a concept that follows the terminology of the traditional factor proportions analysis, but unfortunately the precise meaning of the concept might in this context seem unclear. It is required that the different variants are similar enough so that it is meaningful to aggregate over their output volumes, and hence think of them as an industry.

18. See e.g. Caves and Jones (1977), equation (7.S.16).

19. See Jones (1965), Section IV.

20. See Jones (1968), p. 267.

21. See Jones (1965, equation (11')).

22. Cf. Dixit and Stiglitz (1977), equations (5) and (9).

REFERENCES

Caves, R. and R. Jones, 1977, *World Trade and Payments: An Introduction.* 2nd ed., Little, Brown and Company.

Dixit, A. and N. Norman, 1980, *Theory of International Trade*, Cambridge University Press.

Dixit, A. and J. Stiglitz, 1977, "Monopolistic Competition and Optimum Product Diversity", *American Economic Review*, 67, No. 3, pp. 297-308.

Eaton, J. and H. Kierzkowski, 1983, "Oligopolistic Competition, Product variety and International Trade", in Kierzkowski, H. (ed), *Monopolistic Competition and International Trade*, Oxford University Press, (forthcoming).

Ethier, W., 1979, "Internationally Decreasing Costs and World Trade", *Journal of International Economics*, 9, pp. 1-24.

Hanoch, G., 1975, "The Elasticity of Scale and the Shape of Average Costs", *American Economic Review*, 65, No. 3, pp. 492-497.

Helpman, E., 1981, "International Trade in the Presence of Product Differentiation, Economies of Scale and Monopolistic Competition: A Chamberlin-Heckscher-Ohlin Approach", *Journal of International Economics*, 11, pp. 305-340.

Helpman, E., 1983, "Increasing Returns, Imperfect Markets, and Trade Theory", in R. Jones, and P. Kenen, (eds) *Handbook of International Economics*, North-Holland, (forthcoming).

Jones, R., 1965, "The Structure of Simple General Equilibrium Models", *Journal of Political Economy*, 73, pp. 557-572.

Jones, R., 1968, "Variable Returns to Scale in General Equilibrium Theory", *International Economic Review*, 9, pp. 261-272.

Krugman, P., 1979, "Increasing Returns, Monopolistic Competition, and International Trade", *Journal of International Economics*, 9, pp. 469-480.

Krugman, P., 1980, "Scale Economies, Product Differentiation and the Pattern of Trade", *American Economic Review*, 70, No. 5, pp. 950-959.

Lancaster, K., 1980, "Intra-industry Trade under Perfect Monopolistic Competition", *Journal of International Economics*, 10, pp. 151-175.

CHAPTER III

PRODUCT DIVERSITY, TRADE, AND WELFARE

CHAPTER III PRODUCT DIVERSITY, TRADE, AND WELFARE*

1. INTRODUCTION

Does a market economy yield brand proliferation in excess of what is socially desirable? Or, put differently, is the profit signal appropriate in establishing the socially optimal diversification and output of each product? This question has been debated for at least half a century. Chamberlin (1933), for instance, argued that firms will operate with excess capacity, and that there will be excessive product differentiation in a monopolistically competitive market equilibrium, as compared to a perfectly competitive one. This was conventional wisdom for 50 years, although some economists, e.g. Hotelling (1929), expressed other views. In the mid-seventies two influencial papers, Spence (1976), and Dixit and Stiglitz (1977), challenged the wisdom of this belief, and argued that the market equilibrium instead exhibits too *little* diversity, relative to a first-best social optimum.[1] These two papers also considered ways of regulating the market equilibrium when the first-best policy isn't available, and examined some properties of these second-best equilibria. Dixit and Stiglitz (1977) concluded that a second best optimum where each firm's output volume is regulated, is identical to the market equilibrium. Spence (1976) also showed this, and in addition found that this type of regulation is less efficient in raising social welfare than one that regulates the number of firms.

Koenker and Perry (1981) presented a framework within which some new light is shed on the questions mentioned above. Their answers to these questions depend, in particular, on the degree to which there are increasing returns to scale, and on firms' conjectures about rivals' behavior.

The present chapter considers these problems in a framework similar to that of Koenker and Perry (1981) but from another point of view: It analyzes whether changes in the *size* of the economy, as measured by its factor endowment, affect the relation between the market equilibrium, and first - or second - best optima. It will be shown that size indeed is crucial to the probability of excess diversity and the relative efficiency of different second - best policies, as long as the perceived elasticity of demand for a differentiated good is not confined to being constant. This change in the scale of factor endowments is interpreted as the result of the opening up of international trade between economies that are scaled replices of each other. An alternative view would be a growing, closed economy interpretation where we compare different steady states. A by-product of our analysis is that we will analytically derive some of the results that Koenker and Perry (1981) obtained with numerical simulation techniques.

Our analysis is based on a one-factor, two-sector general equilibrium model, with a demand side representation that is by now common, particularly in the theory of international trade in differentiated products. We express our variables of interest as functions of the factor endowment and compare different equilibria for various sizes of this endowment. Following Koenker and Perry (1981), we make different assumptions about the individual firm's conjectures regarding competitors' strategies, and about its share of the market, compared to the authors referred to above. Chamberlin (1933), as well as Dixit and Stiglitz (1977), assume that firms behave in a Cournot or Bertrand fashion, and that the individual firm is of negligible size relative to the market. Spence (1976) assumes that each firm expects the rest of the industry to accommodate any marginal change the firm undertakes in its supply. Here, on the other hand, we do not restrict conjectures to either of these types, but allow in this respect for a more general treatment. The important difference is that the individual firm's perceived elasticity of demand is no longer constant[2].

The model and the resulting equilibria are characterized by strong symmetry due to particular assumptions that are made. However, these special assumptions simplify the analysis considerably and they are hopefully justified on the ground that we are not presently interested in biases in product selection[3]. But it deserves to be emphasized that by assuming that consumers have identical preferences many essential aspects of product differentiation are not captured[4]. We believe, however, that our analysis deals with phenomena that are much more general than our special assumptions would seem to imply.

Terms like "planner" and "regulation" are used in the literature referred to above, and are also employed here. But we like to stress firmly that we are completely neglecting *how* the regulation should be undertaken and the associated *costs*. No serious policy-oriented welfare analysis of regulation could disregard these aspects.

Our analytical framework stands in the following relation to the previous literature. Just as in Dixit and Stiglitz (1977), we rely on a two-sector general equilibrium model, as opposed to Spence's (1976), and Koenker's and Perry's (1981), partial equilibrium analysis with neglected income effects. Dixit and Stiglitz (1977), and Koenker and Perry (1981), assume a fixed cost and a constant marginal cost technology, whereas in Spence (1976), and also here, there are more general representations of the technology in the industry of particular interest. Preferences are more general in Spence (1976), and in Dixit and Stiglitz (1977), than they are here or in Koenker and Perry (1981). Conjectures about competitors' responses are the same here as in Koenker and Perry (1981); they include the other two papers assumptions as special cases. In other words, on the production side the assumptions are equally or less restrictive, and on the demand side equally or more restrictive, than in the other three studies.

The chapter is structured as follows. In the next section the model is presented in detail, and demand functions are derived, as well as the equations that define the market equilibrium. The following section 3 solves the problem facing the planner of finding the first best social optimum. The market equilibrium and the unrestricted social optimum are compared in section 4 for various sizes of the economy. Section 5 compares two non-first best regulatory policies. Finally, section 6 offers a summary, and some concluding remarks.

2. THE MARKET EQUILIBRIUM

Envision an economy which produces goods in two sectors, X and Y, using the same mobile factor of production L, which is referred to as "labor". One sector, industry Y, is supposed to represent "the rest of the economy".[5] In this crude characterization it produces a homogeneous good under constant returns to scale, and the good is sold on a perfectly competitive market. The condition for profit maximization and industry equilibrium in sector Y is therefore simply that price equals unit cost, i.e.:

$$p_y = a_y w, \qquad (1)$$

where p_y and w are the prices of the output and of labor, respectively, and where a_y is the (constant) requirement of labor per unit of output. In what follows w is the numeraire, set at unity. We define furthermore, for convenience, units so that $p_y = 1$.

The industry of prime interest is industry X. It consists of n firms, each producing only one variant of good X, and each being the sole producer of that particular specification.[6] Due to product differentiation on the demand side, each producer enjoys some monopoly power. There are economies of scale at the firm level, and hence there will only be a limited number of firms active in a zero-profit equilibrium, where each firm maximizes its profit given its conjectures about competitors' strategies.[7,8]

The technology of any firm i ($i = 1, \ldots, n$) can be represented by the same cost function f, which for each output level x_i gives the minimum requirement of labor ℓ_i. We assume that the economies of scale are eventually exhausted, at some finite output volume. Let us define a measure of the degree of economies of scale as:

$$\theta(x_i) \equiv \frac{f'(x_i)}{f(x_i)/x_i}\ , \tag{2}$$

i.e. marginal over average cost. Our assumptions about the returns to scale can then more exactly be stated as:

$$0 < \theta(x_i) < 1 \quad \text{for } 0 < x_i < \bar{x}\ , \tag{3a}$$

$$\theta(x_i) \geq 1 \qquad \text{for } x_i \geq \bar{x}\ , \text{ and} \tag{3b}$$

$$\theta'(x_i) > 0 \qquad \text{for } x_i > 0\ , \tag{3c}$$

for some finite $\bar{x}$.

Firm i's profit is:

$$\pi_i = p_i x_i - f(x_i), \tag{4}$$

and the first-order condition for profit maximization with respect to output is:

$$p_i(1 - \frac{1}{\varepsilon_i}) = f'(x_i), \tag{5}$$

where ε_i is the price elasticity of demand for good i.[9]

Of prime interest in the subsequent analysis are the properties of the elasticity of demand ε_i. To derive an expression for this elasticity we have to specify the demand side. We will assume that the representative consumer's preferences can be represented by a two-stage Cobb-Douglas-CES utility function:

$$u = \left(\sum_{j=1}^{n} x_j^{\beta}\right)^{\frac{\alpha}{\beta}} Y^{(1-\alpha)} \text{ where } 0 < \alpha < 1 \text{ and } 0 < \beta < 1, \tag{6}$$

and where Y denotes the amount of good Y consumed and n is the number of products in existence. This representation resembles the general equilibrium representations used by Dixit and Stiglitz (1977), although they alternatively allowed for a more general formulation in the upper or lower stages, and the partial equilibrium representations employed by Spence (1976), and by Koenker and Perry (1981). In standard fashion, the consumer maximizes utility subject to a budget constraint. If I is total income the demand functions become:

$$p_i = \frac{x_i^{\beta-1}}{\sum_j x_j^\beta} \alpha I \text{ and } p_y = \frac{(1-\alpha)I}{Y}, \tag{7, 8}$$

when written on implicit form.

Let us now revert to the elasticity of demand as perceived of by a typical firm i. From (7) we derive:

$$-\frac{1}{\varepsilon_i} = \beta - 1 - \frac{x_i}{z}\frac{dz_i}{dx_i}, \tag{9}$$

where $z \equiv \sum_{i=1}^{n} x_i^\beta$, and where

$$\frac{x_i}{z}\frac{dz_i}{dx_i} = \frac{\beta x_i}{z}\left(x_i^{\beta-1} + \sum_{j \neq i} x_j^{\beta-1}\frac{dx_j}{dx_i}\right). \tag{10}$$

The elasticity of demand for firm i can, as equation (9) shows, be thought of as consisting of two parts. One is the direct effect when the industry's total output volume (as measured by z) remains unchanged, and the second is the effect of the induced change in this volume. In Spence (1976), and in Dixit and Stiglitz (1977), it is assumed that the latter effect - the expression in (10) - is equal to zero[10,11]. The implication of this is of course that the elasticity of demand is equal to $\frac{1}{1-\beta}$, i.e. it is invariant both to the number of competing products and to the amount in which the different varieties are consumed. This assumption obviously facilitates the ensuing comparative statics, and might therefore be defended, if the question at issue so permits. It might also be justified if one believes that it is actually a good description of the industry's product space; i.e. that the entry of new varieties does not make it more "crowded". This, of course, is contradictory to the basic idea of locational models of product choice and, we would argue, to our usual perceptions of the product spaces of most differentiated products.[12]

Our purpose is to analyze not only economies where n is very large (as studied by Dixit and Stiglitz [1977]), but also those in which there are fewer firms. Since Cournot conjectures are often argued to be unreasonable for smaller markets, it seems desirable to allow for other types of behavior.

We will, following Koenker and Perry (1981), assume that each firm's conjectural derivative can be approximated by:

$$\frac{dx_j}{dx_i} = \frac{\delta}{n-1}, \ \forall j \neq i, \tag{11}$$

with $\delta \geq -1$. $\sum_{j \neq i} \frac{dx_j}{dx_i} = \delta$ is hence the total response by all other firms conjectured by the oligopolist, which possibly is a function of the number of firms in the industry. To facilitate the subsequent analysis we assume that $(1+\delta)/n$ goes to zero as n goes to infinity. If a firm entertains Cournot conjectures, i.e. no response is expected from the rest of the industry, $\delta = 0$. If the rest of the industry is believed to accommodate any change in firm i's quantity, $\delta = -1$. All the subsequent results are qualitatively similar for all values of $\delta > -1$; Cournot conjectures are hence included as a special case.

Since the assumptions we have made imply complete symmetry on both sides of the market, a potential equilibrium is a symmetric one where all firms produce the same quantity x, and sell at the same price p. All the different types of equilibria that are treated below are symmetric in this way, and the symmetry allows us to rewrite equation (9) for the representative firm as:

$$-\frac{1}{\varepsilon} = \beta\left(1 - \frac{(1+\delta)}{n}\right) - 1. \qquad (12)$$

The first-order condition for profit maximization is therefore:

$$p\beta\left(1 - \frac{(1+\delta)}{n}\right) = f'(x). \qquad (13)$$

For marginal revenue to be positive it is required that $n > \delta + 1$, which is henceforth assumed. As usual we also have for the market equilibrium an industry equilibrium condition which states that profits are zero:

$$px = f(x). \qquad (14)$$

Finally we require that labor and goods markets are in equilibrium, respectively:[13]

$$nf(x) + Y = L, \tag{15}$$

$$pnx = \alpha I, \tag{16}$$

and

$$Y = (1 - \alpha) I. \tag{17}$$

We are now able to characterize the market equilibrium E. Equations (13)-(17) constitute, due to Walras' Law, four independent equations that suffice to determine n, x, Y, p. Equations (13) and (14) imply

$$\beta\left(1 - \frac{1 + \delta}{n^e}\right) = \theta(x^e), \tag{18}$$

where superscript e denotes market equilibrium values. Since we have assumed that θ is a monotonic function of x, it is invertible. Our quantities of primary interest are then:

$$x^e = \theta^{-1}\left[\beta\left(1 - \frac{(1 + \beta)}{n^e}\right)\right], \tag{19}$$

and, due to (15) and (17),

$$n^e = \frac{\alpha I}{f(x^e)} . \tag{20a}$$

But since we have zero profits in both sectors in E, all income is factor income, and therefore:

$$n^e = \frac{\alpha L}{f(x^e)} . \tag{20}$$

Subsequently we will refer to L as the "size" of the economy. We assume that households are identical in all respects, and that the supply of labor is fixed for each. Increasing the size of the economy is hence equivalent to increasing the number of workers-households. It is clear from equations (19) and (20) that both x^e and n^e are affected by the size of L.

To interpret the increase in size as the opening up of frictionless international trade shouldn't cause any problem, since with identical technologies and preferences, factor price equalization will occur.[14] Due to the symmetry of the demand side, each producer faces the same elasticity of demand, regardless of the location of production. With n^e in equation (19) now denoting the number of products in the world, each producer chooses the same total output x^e, and therefore receives the same price p^e. The zero profit conditions then assure that factor prices equalize, and since the trading economies are scaled replicas of each other there will be no net trade, although there is intra-industry trade in differentiated products.[15]

Let us explore the relation between the equilibrium output volume per firm and the size of the economy. Industry and firm equilibrium requires that $\theta(x^e) = 1 - \frac{1}{\varepsilon}$. The maximum value that ε can take on is $\frac{1}{1-\beta}$, which occurs when n^e is infinitely large.[16] Therefore the maximum value of $\theta(x^e)$ is β, and as $\beta \leq 1$ no firm will ever operate "to the right" of its minimum avarage cost. Now, if we use equation (20) to substitute for n^e in equation (18) we get

$$\frac{\beta - \theta(x^e)}{f(x^e)} = \frac{\beta(1 + \delta)}{\alpha L} , \tag{21}$$

and hence, since average cost-minimum by assumption is at a finite x^e,

$$x^e \to \theta^{-1}(\beta) \quad \text{as } L \to \infty \tag{22}$$

3. THE UNRESTRICTED SOCIAL OPTIMUM

In order to determine whether or not there is excess diversity in the market equilibrium, we obviously need to know what the optimal allocation is. It is as usual possible to think of several different concepts of optimality, depending on how the set of feasible allocations is restricted. In this section we will consider what will be referred to as an *unrestricted* social optimum U. This is the allocation that a planner who maximizes the representative consumer's utility B, subject only to given preferences and technology would choose. In the international trade interpretation the planner hence decides on the optimal world allocation. The optimum is unrestricted in the sense that the planner is free to set both the number of firms, and their output volumes. We will in the next section also consider a couple of optima where the planner is constrained to the use of just one of these instruments.

The symmetry of preferences and technology allows us to formulate the planner's problem as:

$$\max_{n,x,Y} \quad B = n^{\frac{\alpha}{\beta}} x^{\alpha} Y^{1-\alpha}, \tag{23}$$

$$\text{s.t.} \begin{cases} \ell = f(x), \text{ and} & (24a) \\ n\,\ell + Y = L, & (24b) \end{cases}$$

or

$$\max_{n,x} \quad B = n^{\frac{\alpha}{\beta}} x\alpha\Big(L - nf(x)\Big)^{1-\alpha}. \tag{25}$$

First-order conditions are (letting superscript u denote the unrestricted optimum):[17]

$$n^u = \frac{\alpha L}{f(x^u)} \left[(1 - \alpha)\beta + \alpha\right]^{-1}, \tag{26}$$

and, when we take (26) into account,

$$x^u = \theta^{-1}(\beta). \tag{27}$$

This result, that the optimal firm size is independent of the size of the economy, can be interpreted as follows. The marginal rate of substitution in consumption between n and x is $\beta n/x$, whereas the marginal rate of transformation between the two is $\theta(x)n/x$. Hence, due to the economies of scale the MRT changes at a different rate than the MRS when x changes, and therefore only one particular x can be optimal.

4. A COMPARISON OF THE MARKET EQUILIBRIUM AND THE SOCIAL OPTIMUM

We are now in a position to compare the allocation in the market equilibrium with that in the social optimum. It is immediate from equations (19) and (27) that, since θ is in creasing in x and since $0 \leq \frac{1+\delta}{n} < 1$, it will generally be the case that:

$$x^e < x^u, \tag{28}$$

that is, firm size in the market equilibrium will be less than what is socially optimal. The result that these are equal, obtained by Spence (1976), and by Dixit and Stiglitz (1977), is hence a special case that arises when $\frac{1+\delta}{n} = 0$. It is also clear that if the representative firm believes that its competitors seek a market share solution, the difference between x^e and x^u would be at a maximum[18].

Let us now turn to the question of diversity. Equations (20), (26), and (27) give

$$\frac{n^e}{n^u} = \frac{f(\theta^{-1}(\beta))[1 - \alpha)\beta + \alpha]}{f(x^e)} \tag{29}$$

We first note that, for x^e sufficiently close to, though not necessarily equal to, $\theta^{-1}(\beta)$, the ratio on the RHS is less than unity. Since x^e converges towards x^u when L grows infinitely, it seems that Spence's (1976), and Dixit and Stiglitz's (1977), assertion that there is too little diversity in the market equilibrium, is most likely to be valid for the very large economy. On the other hand, with x^e small enough, i.e. for the sufficiently small economy the opposite will prevail, at least if the fixed-cost element is "small".[19] That is, it is also possible that the market equilibrium implies an excess diversity relative to the social optimum. This possibility was demonstrated by Koenker and Perry (1981) in their numerical analysis, and by Lancaster (1980). What is shown here is that, given the degree of product differentiation and the degree of economies of scale, and given firms' conjectures about competitors' strategies, *excess diversity in the market equilibrium is more likely the smaller the economy's factor endowment*, and that *it is the large economy that is likely to have too little diversity.* Furthermore, *there exists some finite size $\tilde{L}$ of the economy,* *for which the number of firms in the market equilibrium is equal to that in the first best optimum.*

Not only the market equilibrium, but also the reference point - the unrestricted social optimum - is influenced by the factor endowment. We will consider the relation between utility levels in the two allocations in more detail below. But it is clear that there is a gain from trade that derives from the increase in product variety it brings about, regardless of whether or not a country's autarky factor endowment is smaller or larger than $\tilde{L}$. It is possible that consumers in a country that face excess diversity in autarky, will after trade is opened up, find themselves living in a world with more, but now too little, diversity, in both cases relative to the respective first-best optimum. Trade may in this way actually *increase* the relative divergence between the market equilibrium and the social optimum in terms of product variety.

One way of interpreting these results is the following. Each firm's perceived elasticity of demand increases as the economy becomes larger through increased competition from entering firms. With a fixed budget share for differentiated goods, and an increasing output of each, the number of product varieties must increase proportionally *less* than the size of the economy, i.e. n^e is strictly concave in L. Two conflicting forces influence whether there is excess diversity in the market equilibrium. One is that the X-industry uses more resources in the social optimum than in the market equilibrium. This is clear from a comparison of equations (20) and (26), which we restate:

$$n^e f(x^e) = \alpha L \tag{20}$$

$$n^u f(x^u) = \left[\alpha(1 - \beta) + \beta\right]^{-1} \alpha L \ , \tag{26}$$

where $[\alpha(1 - \beta) + \beta]^{-1} > 1$. This tends to make $n^u > n^e$. On the other hand, less resources are devoted to the production of each variant in the market equilibrium than in the social optimum, since $x^e < x^u$. Since these resources can be used to produce additional specifications, there is also a tendency for $n^e > n^u$. The smaller the economy, the lower the equilibrium elasticity of demand, and the larger the difference between x^e and x^u. Hence for L sufficiently small n^e exceeds n^u, whereas the opposite prevails when L is sufficiently large, and at $\tilde{L}$ these forces are just balancing each other.

The market equilibrium E, and the unrestricted social optimum U are shown in Fig. 1 in (n , x)-space. ZP is the zero profit locus and $PM(\delta_1)$ and $PM(\delta_2)$ are two profit maximizing loci where $\delta_2 > \delta_1$. The "elliptic" contours around U are iso-benefit contours. Whether n^u is larger or smaller than n^e obviously depends on the size of δ. But, as depicted in Fig. 2 and 3, country size also matters.

Consider next consumer utility levels. Per capita utility is higher the larger the economy regardless of whether it is a market economy or not. This gain from trade stems from the fact that in the larger economy each consumer can consume less of each specification, but more different types of goods, *without* incurring any higher average costs. Per capita utility in the market economy is of course in most cases lower than in the social optimum. But

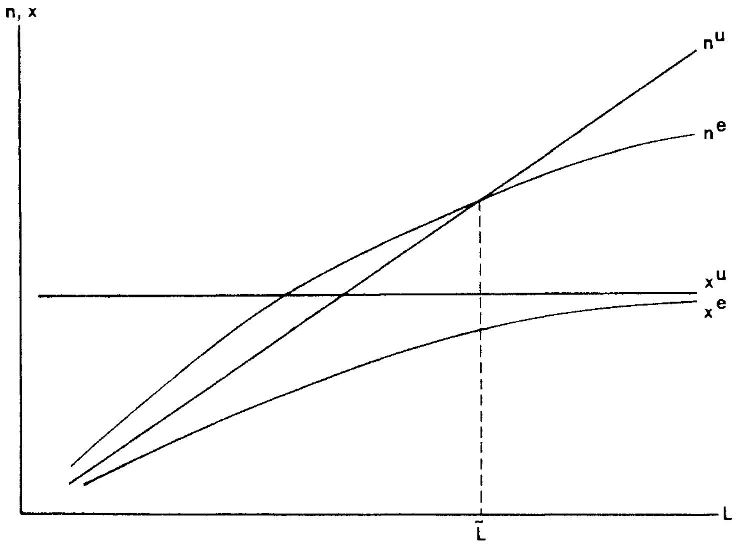

Figure 1

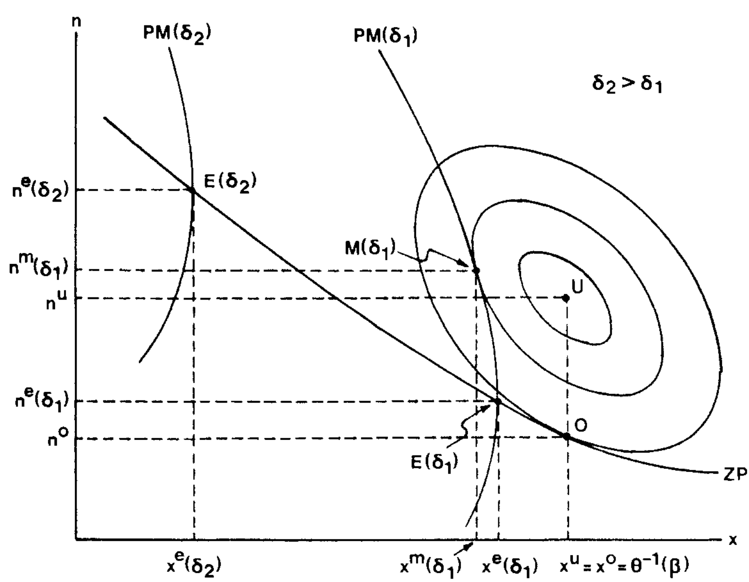

Figure 2

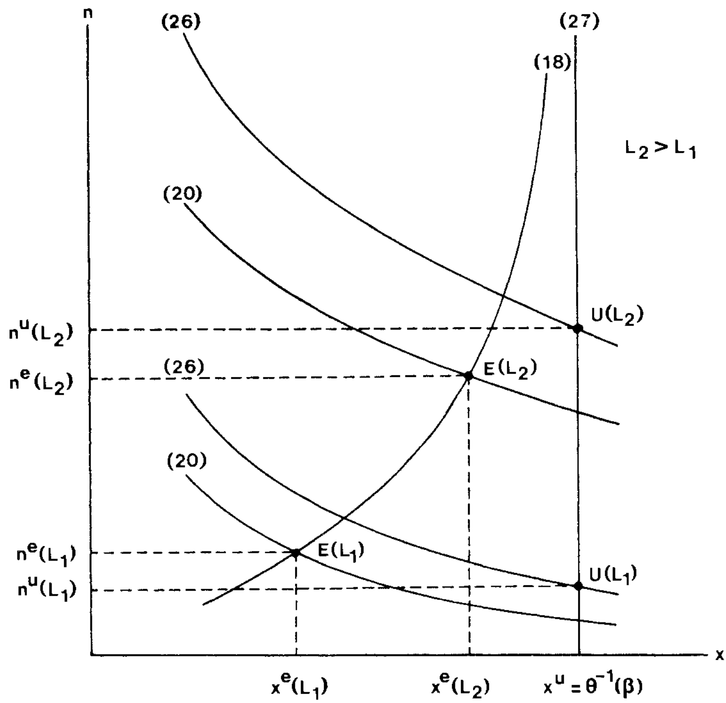

Figure 3

what happens when the economy grows infinitely large? As we saw above $x^e \to x^u$, and hence we have

$$\frac{(B^e/L)}{(B^u/L)} \to \frac{[\alpha(1-\beta)+\beta]^{(\frac{\alpha}{\beta}+1-\alpha)}}{\beta^{1-\alpha}} \qquad \text{when } L \to \infty. \tag{30}$$

The limit value is less than unity for β less than unity, and therefore, *the per capita utility level in the market equilibrium will generally not converge towards the level in the corresponding social optimum.* But, it is shown in the appendix that the RHS is increasing in β, and hence *the less differentiated products are, the less will be the difference between the two per capita utility levels be.* The opening up of international trade thus in this sense reduces, even though it usually doesn't completely eliminate, the scope for regulation of the market economy.

5. A COMPARISON OF TWO CONSTRAINED OPTIMA

In the unrestricted social optimum treated above, the planner was free to choose both the representative firm's output volume and the number of producers in the differentiated goods industry. Let us now assume that our planner is less omnipotent in that just one of these instruments is available at a time. Consider first a regulatory policy that aims at controlling the *size* of the representative firm, while the profit potential determines the number of firms active in the market, just as in the market equilibrium. This can be called a *behavioral regulation* since we don't allow the firms to act in the way they would otherwise do. The planner's problem is hence to choose only x and leave n free to vary, so as to maximize the benefit function (25) subject to the zero profit condition (14) and the market demand function (16). Since free entry and exit will eliminate all profits, factor services will constitute the only source of income, just as in the market equilibrium. The solution to the maximization problem is the constrained equilibrium O, with the following solution:[20]

$$x^o = \theta^{-1}(\beta) \qquad (31)$$

$$n^o = \frac{\alpha L}{f(x^o)} \quad . \tag{32}$$

As demonstrated also by other authors the planner will choose the same output volume when restricted to the behavioral regulation as would prevail in the unrestricted social optimum. It is immediate from comparing equation (32) to the corresponding equation (20) and (26), that the number of firms in the social optimum that is restricted to behavioral regulation is *less* than *both* what is socially fully optimal, *and* what will prevail in the market equilibrium. But the welfare level is of course in between these two allocations.

A comparison of equations (31)-(32) with (19)-(20) shows why Dixit and Stiglitz (1977) had a case where the market equilibrium is identical to the social optimum where only output per firm can be regulated: they assume that $\frac{1+\delta}{n} = 0$, implying $x^o = x^e$ and $n^o = n^e$. Therefore, once again, the results of Dixit and Stiglitz (1977) are valid for the very large economy, but not necessarily for smaller economies.

The restricted social optimum O is shown in Fig. 1. Due to the zero profit restriction it obviously has to be located along the ZP locus, at the point of tangency with the highest iso-benefit contour.

Let us now consider whether the efficiency of the behavioral regulation in increasing welfare is dependent on the size of the economy. If we measure the efficiency of the regulatory policy by the relative increase in welfare it brings about, and if we take the market equilibrium as a basis, our measure is:

$$\mu(L) \equiv \frac{B^o(L) - B^e(L)}{B^e(L)} = \frac{B^o(L)}{B^e(L)} - 1. \tag{33}$$

From the expressions above we have that:

$$\frac{B^o(L)}{B^e(L)} = \left[\frac{f(x^o)^{-\frac{1}{\beta}} x^o}{f(x^e)^{-\frac{1}{\beta}} x^e}\right]^{\alpha}. \tag{34}$$

Since x^o is determined by β, and since

$$\frac{d}{dx^e}\,(f(x^e)\; x^e) > 0 \quad \text{and} \quad \frac{dx^e(L)}{dL} > 0, \tag{35}$$

B^o/B^e must be decreasing in L. This implies that:

$$\frac{d\;\mu(L)}{dL} < 0, \tag{36}$$

i.e., *the behavioral regulation is in relative terms more efficient the smaller the economy*. This is intuitively appealing when we recall the result that the output quantity which is established in the market equilibrium approaches what is socially desirable as the economy grows. Hence, it seems natural that the scope for this type of policy shrinks as the economy gets larger.

Another type of regulatory policy is *structural* regulation, which aims at controlling the *number* of firms in the market, while leaving to the individual firm to set its quantity at a profit maximizing level. The planner's problem is hence to maximiz benefits (25), subject to the first-order condition for profit maximization (13), and the market demand function (16). The solution is depicted graphically as the point M in Fig. 1, for the case where $\delta = \delta_1$. An algebraic solution in the form presented above is hard to derive since the constraint is an implicit function on a somewhat complicated form. Nevertheless, some conclusions can be drawn concerning the efficiency of this regulatory policy. We recall that for low values of L, we have $n^e > n^u$, whereas for L larger than $\tilde{L}$ we have $n^e < n^u$ (see Fig. 2). It is therefore immediate that *at $\tilde{L}$ there is nothing to be gained from structural regulation.* But *the more the size of the economy differs from $\tilde{L}$ the more scope there is for this policy.*

A direct comparison of the efficiency of the two policies for various L is difficult as we have not derived the optimal n and x for the structural regulation. But as the structural policy is pointless at $\tilde{L}$, the behavioral policy must clearly be superior for the size. Since the market equilibrium output volume converges towards the social first best volume when L grows larger than $\tilde{L}$, and simultaneously the number of firms in the market equilibrium falls relative to what is socially optimal, we can make the following statements;
There exists a $\bar{L} > \tilde{L}$ such that for $L > \bar{L}$ structural regulation is most fruitful. There will also be a $\underline{L} < \tilde{L}$ such that for $\underline{L} \leq L \leq \bar{L}$ the behavioral regulation must yield a higher benefit level. For $L < \underline{L}$ it is indeterminate which policy performs best.[21]

6. SUMMARY AND CONCLUDING REMARKS

We have shown that, as the size of the economy increases through the opening up of international trade, or through growth in the factor endowment,

(a) output per firm in the market equilibrium approaches the first best optimum;

(b) the number of firms in the market equilibrium changes gradually from being in excess to being too few, as compared to the unrestricted social optimum;

(c) the scope for a regulatory policy that aims at controlling the firms' behavior decreases;

(d) the welfare improving potential of an entry controlling policy is first reduced and thereafter increased: and that

(e) welfare in the market equilibrium will generally not approach the welfare level of the unrestricted social optimum, but will, in the limit, converge towards a level that is negatively correlated with the degree of product differentiation.

Finally, due to (c) and (d), when we compare the efficiency of the two regulatory policies

(f) all possible sizes of the economy can be devided into three mutually exclusive intervals. In the highest, structural regulation is most efficient, in the middle interval behavioral regulation is preferred, whereas for the lowest region the result is ambiguous.

A conceptual difference between this paper and most of the earlier literature is that we allow for situations where firms believe that their actions will influence the actions of others; that is, where the perceived cross elasticity of demand is different from that when firms hold Cournot conjectures, and where firms' market shares are not close to zero.

That firms take into account how their actions affect others makes the framework suitable for the analysis of other cases than Chamberlin's (1933), and Dixit's and Stiglitz's (1977), "large group" industry. It seems that their models are most suitable for the analysis of large economies. On the one hand, they assume that there are *many* firms active in the industry, so as to motivate e.g. Bertrand conjectures. On the other they require that the industry is well-defined, and possible to distinguish from the rest of the economy. The second point seems to require that the elasticity of substitution between different variant of the differentiated good is high, to make products similar from the consumers' point of view. But the higher this elasticity, the larger the size of the representative firm, and hence the *lower* the number of firms in equilibrium. The number of firms will be larger again if the industry constitutes a large proportion of the economy, but this might run contrary to our attemt at making a clear distinction between the industry and the rest of the economy. A way out is hence to assume that the economy is large. For cases other than this, however, the framwork presented here should have some merit.

FOOTNOTES

* Thanks are due to James Markusen, Torsten Perssson, and Lars E.O. Svensson for valuable discussions. They also, together with Avinash Dixit, Wilfred Ethier, Harry Flam, Gene Grossman and Carl Hamilton, provided comments on earlier drafts, as did participants in seminars in the Stockholm Theory Workshop, and at the Institute for International Economic Studies. I am also grateful to Humanistisk Samhällsvetenskapliga Forskningsrådet for financial support.

1. We refer to the sections of the two papers where a CES utility index for the differentiated goods is employed.

2. Spence (1976), and Dixit and Stiglitz (1977), also consider cases where the elasticity of demand is not confined to being constant, but then this is due to other formulations of the preferences over the differentiated goods, rather than this supply side phenomenon.

3. Such biases are studied in Spence (1976).

4. Locational models as those used by Hotelling (1929), and Lancaster (1979), attempt to capture some of these aspects.

5. L might be thought of as the total endowment of good Y, $\bar{Y}$ the amount consumed and $L-\bar{Y}$ the amount that is transformed into X goods. In particular, as suggested by Dixit and Stiglitz, we can interpret the economy's factor endowment as the time at the disposal of the consumers. If each individual has a fixed amount of time available, L will be the number of consumers, and good Y might be interpreted as leisure.

6. There are generally, in a model such as this, incentives for firms to gain control over other firms in order to co-ordinate output decisions, but this is not dealt with here.

7. We need to assume, strictly speaking, that the marginal cost exceeds the marginal revenue at $x_i = 0$. This would be taken care of by any fixed cost, since the marginal cost would then be infinite at this point.

8. All cooperative solutions are assumed away.

9. The second-order condition is that the elasticity of marginal cost exceeds that of marginal revenue, i.e.:

$$-\frac{1}{\varepsilon_i} + \frac{1}{(\varepsilon_i - 1)} \frac{x_i}{\varepsilon_i} \frac{d\varepsilon_i}{dx_i} < \frac{x_i f''(x_i)}{f'(x_i)} \equiv \tau(x_i).$$

$\tau(x_i)$ and $\theta(x_i)$ are related in the following way

$$\frac{x_i \theta'(x_i)}{\theta(x_i)} = 1 - \theta(x_i) + \tau(x_i).$$

10. However, cf. footnote 1.

11. Dixit and Norman (1980) motivate this with, first, that firms hold Cournot conjectures and hence $dx_j/dx_i = 0$ for all $j \neq i$, and second, that the firm under study is assumed to be small relative to the market and therefore $\beta x_i^\beta / \Sigma_j \, x_j^\beta$ is approximately zero. The same approximation is done in Dixit and Stiglitz (1977), but with Bertrand type conjectures.

12. For a discussion of this issue see Pettengill (1979), and Dixit and Stiglitz (1979).

13. n is henceforth treated as a real number, although it should be an integer.

14. See e.g. Dixit and Norman (1980), or Helpman (1981), for an explicit treatment of the international equilibrium.

15. Net trade is the net exchange of differentiated goods for food.

16. Lancaster (1980) calculates an inter-group elasticity of demand with this property, using the same demand representation.

17. It is shown in the appendix that a sufficient, although not necessary, condition for B to have a local maximum at (n^u, x^u) is that $x^u f''(x^u)/f'(x^u) > -(1-\beta)$; so the marginal cost might hence be decreasing in the output volume at the equilibrium point.

18. That is, each individual firm strives to keep its share of the market, implying $dx_j/dx_i = x_j/x_i = \delta/n-1$, or by symmetry that $\delta = n - 1$. As we assumed that $n > \delta + 1$ above, we hence only allow conjectures to be *close* to that of a market share behavior. (This restriction wouldn't be necessary with non-constant budget shares for the two types of goods).

19. Fixed costs must be small to allow for a sufficiently small $f(x^e)$.

20. A sufficient second-order condition is that $\eta = 0$.

21. We can distinguish additional unambiguous intervals, if we allow ourselves to completely ignore the integer problem.

if $L < \tilde{L}$, and $L \to 0+$, then $n^e \to n^u+$, whereas x^e diverges more and more from x^u. There will therefore be an additional interval $(0, \underline{\underline{L}})$ in which behavioral regulation is more efficient. We hence have two possibilities: Either $\underline{\underline{L}} > \underline{L}$, in which case behavioral regulation is preferable for all $L \leq \bar{L}$, whereas structural regulation has a larger welfare improving potential for $L > \bar{L}$. Alternatively, $\underline{\underline{L}} < \underline{L}$, in which case the sequence of preferable second-best policies is, starting at $L = 0$: behavioral, structural, behavioral, structural.

APPENDIX

$\tau(x^u) \geq 0$ is Sufficient to Fulfill the Second-order Condition

For $B(n,x) = n^{\frac{\alpha}{\beta}} x^{\alpha} \left(L - nf(x)\right)^{1-\alpha}$ we have that

$$\frac{\partial^2 B}{\partial n^2} = - B(\cdot)\left[\frac{\alpha}{\beta}\frac{1}{n^2} + (1 - \alpha)\frac{f(x)^2}{(L - nf(x))^2}\right] \qquad (A1)$$

$$\frac{\partial^2 B}{\partial x^2} = - B(\cdot)\frac{\alpha}{x^2}\left[\frac{1}{1 - \alpha} + \tau\right] \qquad (A2)$$

$$\frac{\partial^2 B}{\partial x \partial n} = - B(\cdot)\frac{\alpha}{xn}\left(1 + \frac{\alpha}{\beta(1 - \alpha)}\right. \qquad (A3)$$

A sufficient condition for the function to have a local maximum at (x^u, n^u) is that the Hessian is negatively definite at this point, i.e.

$$\frac{\partial^2 B}{\partial n^2}\frac{\partial^2 B}{\partial x^2} - \left[\frac{\partial^2 B}{\partial x \partial n}\right]^2 > 0. \qquad (A4)$$

(A1) - (A3) inserted in (A4) gives after some manipulations the condition

$$\tau(x^u) > - (1 - \beta). \qquad (A5)$$

The Limit Value of $(B^e/L)/(B^u/L)$

If we take the logarithm of the limit value in (30) we get

$$(\frac{\alpha}{\beta} + 1 - \alpha)\ \ln\left(\alpha(1 - \beta) + \beta\right) - (1 - \alpha)\ \ln \beta.$$

Differentiating w.r.t. β yields that

$$-\frac{\alpha}{\beta^2}\ \ln\left(\alpha(1 - \beta) + \beta\right) + (\frac{\alpha}{\beta} + 1 - \alpha)\ \frac{1-\alpha}{(\alpha(1 - \beta) + \beta)} - \frac{(1 - \alpha)}{\beta} =$$

$$= -\frac{\alpha}{\beta^2}\ \ln\left(\alpha(1 - \beta) + \beta\right) > 0.$$

The limit value in (30) is hence increasing in β. This implies that the expression in (30) is at its maximum when β approaches unity, in which case

$$\frac{(B^e/L)}{(B^u/L)} \to 1.$$

REFERENCES

Atkinson, A. and J. Stiglitz, 1980, *Lectures on Public Economics*, McGraw-Hill, Maidenhead, Berks.

Chamberlin, E.H., 1933, *The Theory of Monopolistic Competition*, Harvard University Press, Cambridge, MA.

Dixit, A. and N. Norman, 1980, *Theory of International Trade*, Cambridge University Press, England.

Dixit, A. and J. Stiglitz, 1977, "Monopolistic Competition and Optimum Product Diversity", *American Economic Review*, 67, No. 3, pp. 297-308.

Dixit, A. and J. Stiglitz, 1979, "Monopolistic Competition and Optimum Product Diversity: Reply", *American Economic Review*, 69, No. 5, pp. 961-63.

Helpman, E., 1981, "International Trade in the Presence of Product Differentiation, Economies of Scale and Monopolistic Competition: A Chamberlin-Heckscher-Ohlin Approach", *Journal of International Economics*, 11, pp. 305-340.

Hotelling, H., 1929, "Stability in Competition", *Economic Journal*, 34, pp. 41-57.

Koenker, R. and M. Perry, 1981, "Product Differentiation, Monopolistic Competition, and Public Policy", *Bell Journal of Economics*, 12, No. 1, pp. 217-31.

Lancaster, K., 1979, *Variety, Equity, and Efficiency*, Columbia University Press, New York.

Lancaster, K., 1980, "Pettengill versus Dixit-Stiglitz: A Third Party Intervention", Discussion Paper No. 54, Dept of Economics, Columbia University.

Pettengill, J., 1979, "Monopolistic Competition and Optimum Product Diversity: Comment", *American Economic Review*, 69, No. 5, pp. 957-60.

Spence, A.M., 1976, "Product Selection, Fixed Costs, and Monopolistic Competition", *Review of Economic Studies*, 43, No. 2, pp. 217-35.

CHAPTER IV

PRODUCT DIFFERENTIATION AND CUSTOMS UNIONS

CHAPTER IV PRODUCT DIFFERENTIATION AND CUSTOMS UNIONS*, **

The literature on customs unions theory is gargantuan. Nevertheless it strikes us as inadequate, both in the theoretical cases it addresses and in the actual circumstances of the world economy to which it is commonly thought relevant. This paper aims to spell out the details of this inadequacy and to indicate some lines along which the theory ought to be developed.

Section 1 very briefly describes basic ideas in contemporary customs union theory, and section 2 then discusses at some length necessary extensions. A general proposition is advanced in the next section. Section 4 develops a specific model incorporating the new features, and section 5 then exhibits aspects of this model via some special comparative-statics exercises.

1. BASICS OF CUSTOMS UNION THEORY

There is a huge literature in this area for three distinct reasons: (i) the *institutional* fact that there are many types of discriminatory trade besides customs unions proper (preferential trading arrangements, free trade areas, common markets, economic unions, and so on), (ii) the subject demands disaggregation beyond two-country, two-commodity models, and therefore gives rise to *many conceivable combinations* of characteristics of countries, commodities, and trade flows, and (iii) the subject requires comparisons of *distorred*

---

* This chapter was written together with Wilfred Ethier.

equilibria which, in particular in conjunction with the second point, increases the mathematical complexity of the analysis. In practice the latter two factors have proved decisive, with analysis concentrated disproportionately upon the case of a customs union proper (free internal trade combined with a common external tariff).

Two basic ideas stand out from the wealth of specific analysis. The first, which might be called the *Vinerian Description,* is the notion that the net effect of a customs union is dependent on the balance of trade creation and trade diversion. This idea is in truth the very core of customs union theory.

(1) *(Vinerian Description).* Formation of a customs union produces trade creation between the partners, in response to the mutual elimination of tariffs on each other's goods, and trade diversion from third countries to the partners, in response to the tariff discrimination produced by the union.

Trade creation is by presumption beneficial and trade diversion harmful. Union formation is thus seen as a second-best exercise, with one distortion replaced by another. This, together with the need for at least some disaggregation, accounts for the multitude of separate possibilities. Other aspects of union formation have of course also received attention. Notable examples, to which we shall return, include changes in the terms trade-both within the union and vis-a-vis the outside world-and also economies of scale and the degree of competition.

Mention of the terms of trade relative to the rest of the world brings us to the second basic idea, due to Kemp, Vanek, Ohyama and Wan. Setting the extermal tariffs at appropriate levels can cause the member countries, in the aggregate, to trade with the rest of the world exactly the same collection and quantities of goods that, in the aggregate, they traded prior to the union. Then the union will have no effect at all on the rest of the world, and the members must in the aggregate benefit, because the only effect they experience is the freeing of mutual trade. In essence, manipulation of the external tariff schedule allows any customs union to eliminate all trade diversion and so consist only of trade creation. By allowing a small deterioration in its external terms of trade, the union can ensure that also non-members benefit.

> (2) *(Kemp-Vanek-Ohyama-Wan)*. By an appropriate choice of common external tariffs and of lump-sum transfers among members, any customs union can assure that no countries in the world lose, and that some gain, as a result of the formation of the union.

This proposition applies irrespective of initial conditions, and so establishes a sense in which the countries of the world always have incentives to form additional unions until global free trade is achieved.

2. DESIRABLE DEPARTURES

Although the two ideas discussed above hardly begin to do justice to the vastness of customs union theory[1], they do indicate the distinguishing features of the subject and are its most important conclusions. Thus we take the pair as a benchmark from which to inquire in what direction contemporary problems mandate that the subject be further developed. We indicate four directions. The four have by no means been completely ignored by the literature; and in some cases have received formal treatment, but they do depart substantially from the core of received customs unions theory.

(a) *Trade Modification.* The basic element of the conventional theory is tariff discrimination: the *same* good that can be imported free of duty from a partner faces a tariff when it comes from a non-partner country. But partners also trade among themselves some goods which they do not import at all (either before or after the union) from outside countries. For want of a better term, we describe as *trade modification* the change in trade with outside countries due to the elimination of tariffs on goods traded only within the union[2]. Thus trade modification differs from its familiar sister concept of trade diversion in that it can be produced by tariff changes consistent with the Most Favored Nation clause and with the GATT, that is, by economic integration resulting from the nondiscriminatory elimination by the partners of tariffs on goods they import only from each other.

Suppose, as an example, that France and Germany form a customs union, with France eliminating its tariff on German automotive supplies, also initially imported from America. This discriminatory tariff change would presumably induce trade diversion, with imports of German automotive supplies replacing some or all of the imports from America. Now suppose in contrast that France initially imports small cars from Germany and large cars from America, and that the integration takes the form of a French abolition of the tariff on small cars. Such a non-discriminatory tariff change is not usually considered as within the domain of customs union theory, even though it is likely to deflect French demand from American to German automotive products in about the same way as the previous case. Next, suppose instead that there is a unified world automobile industry, with France importing engines from Germany and bodies from America, and that France abolishes its tariff on engines. This again causes trade modification, but now French imports of American bodies will likely rise along with French imports of German engines.

Trade modification differs from trade diversion in that it is not the result of geographical tariff *discrimination* but rather of the replacement of one set of tariffs by another, with the property that the tariffs which are changed apply directly to trade flows between a strict subset of countries. As the example makes clear, degrees of substitutability and complementarity between goods become crucial.

Extending the scope of the theory to encompass trade modification might well therefore open the gates to another flood of special cases - hardly welcome in this field especially. And the bulk of empirical work in the area cannot be interpreted as distinguishing trade diversion from trade modification. But there are good reasons why the theory should include both. First of course is that both are in principle part of the problem. More significant is the large volume of empirical work, by Kravis and Lipsey and many others, that in the last decade has revealed extensive international price discrimination at even quite disaggregated levels. As a result economists now tend, much more than before, to think of international trade as the exchange of differentiated products, rather than of homogeneous goods. This shift in perception calls for a shift in emphasis from trade diversion towards trade modification.

But the most important consideration is the range of phenomena to which the theory of economic integration ought to apply. Although many attempts at customs unions and free trade areas have been made since the Second World War, the only actual, substantially successful, instance to which the theory reflected in the Vinerian Description literally applies is western European integration. An important instance, to be sure, but past history. Customs union theory thus appears seldom to be used as a framwork for thought about contemporary problems. But issues to which a theory of economic intergration *ought* to apply are legion. As a glaring example, consider the dominant post-war exercise in global commercial policy: the cumulative tariff reductions produced by successive GATT rounds. As often emphasized by economists and men of affairs alike, these reductions have concentrated on trade in manufactures among the developed countries and have been much less relevant to the trade of LDCs. Thus we have experienced on an historical scale (and continue to do so as the Tokyo Round cuts are implemented) the very sort of nationally-biased tariff reductions that are the very essence of the theory of economic integration. But our theory has not been thought relevant because those cuts, implemented under the Most Favored Nation clause, directly conflict with the Vinerian Description. The incorporation of trade modification in addition to (or instead of) trade diversion is essential to the practical relevance of customs union theory.

(b) *Intra-Industry Trade and Product Differentiation.* Related to the distinction between trade diversion and trade modification is the observation that customs union theory assumes trade in homogeneous products whereas product differentiation and intra-industry trade are pervasive in actual commerce. Theoretical models of these latter phenomena have been developed in recent years, and extension of these models to the theory of economic integration is obviously called for. Recall that our present widespread appreciation of the large and growing significance of intra-industry trade in the industrial world actually dates from studies of the effects of economic integration in Europe.

(c) *Imperfect Competition and Scale Economies.* These phenomena have been considered in the light of customs union theory in the past. What is clearly called for now is the application of recently developed general equilibrium approaches.

(d) *"Small Tariff Changes.* Customs union theory has generally considered only "large" tariff changes between partners: the complete elimination of internal trade barriers. Exceptions have usually been accompanied by apologetic explanations of the convenience of marginal analysis.[3] Large changes have been considered an essential aspect of the problem because theorists have often had in the backs of their minds the example of European integration, with free internal trade an objective from the start, and because the GATT rules permitting customs unions and free trade areas require the abolition of internal tariffs; mere reciprocal tariff preferences are ruled out. But the latter consideration is eliminated by our interest in trade modification. Furthermore, a concern with contemporary problems should redirect attention toward small changes, rendering marginal analysis of interest in its own right. The successive GATT rounds have, at each stage, constituted an incremental change toward free trade among the developed countries. If translated into tariff equivalents, the numerous Voluntary Export Restraints of the "new protectionism" of recent years likewise constitute incremental changes, though in the direction of economic disintegration. Another potentially important example is furnished by the codes of conduct regarding nontariff barriers formulated during the Tokyo Round. These codes are not automatically binding on the GATT nations, or limited to them, but are to be individually acceded to by individual states. Thus each code seems likely to develop a circle of participating countries applying vis-à-vis each other the provisions of the code as regards the respective nontariff barrier, but applying more restrictive prior standards toward nonparticipating countries. Thus the codes might well produce non-tariff equivalents of preferential trading areas.

It is easier to undertake by stages the departures recommended here. Therefore, in the following section we ignore imperfect competition and scale economies, but allow trade modification and inquire into the consequences of a small preferential change in tariffs, in a conventional context. A general proposition, analogous to results well-known in the theory of trade and domestic distortions, follow easily once attention is directed toward small tariff changes; this propositions is stated in section 3, and is formally proved in the appendix. In the rest of the paper we consider also the remaining two suggested departures from the standard treatment of economic integration: product differentiation, and imperfect competition and scale economies. To high-light the impact of these additions, we construct in section 4 a special formal model that incorporates all four departures, but that abstracts from some features already studied. In section 5 we analyze properties of this model by means of comparative statics experiments. Finally, some brief concluding remarks are given in section 6.

3. AN ADDITIONAL GENERAL PROPERTY

Consider an initial state of discriminatory trade where a group of countries, in a tariff-ridden world, have formed a customs union with free internal trade. Suppose these countries stage a "marginal retreat" from the union by imposing a small tariff on imports from each other. Free initial internal trade implies that a marginal import by any member from a partner yields a welfare increase equal to the welfare loss due to the exports required in payment. Thus the "trade destruction" caused by the marginal internal tariff will entail a zero first-order effect on welfare. But any increased trade with the rest of the world produces a positive first-order effect since the initial external tariffs imply that an additional import increases welfare by more than the sacrifice required to pay for it.

(3) *(Preferential Trade is Better than a Customs Union).* If any group of countries has free mutual trade in a tariff ridden world, their joint welfare will be increased by a marginal tariff on each other's goods that raises total trade with the rest of the world (in the specific sense that the new volume of trade yields the partners jointly more revenue than the old, at the original tariffs and prices), if the prices at which the partners trade with external countries do not change.

In essence, the net benefits form "undoing" some trade diversion and trade modification necessarily swamp the losses entailed by "trade destruction", for a small retreat from a customs union. This partially undercuts the consistent motives for union formation that seem to be supplied by idea (2). Also (3) imply a motivation for tariff preferences that fall short of a complete union. This is quite consistent with the prominent contemporary issues that the previous section argued ought to be addressable by the theory of economic integration.

4. SCALE ECONOMIES AND PRODUCT DIFFERENTIATION: A SPECIAL MODEL

The previous section focused on trade modification as well as trade diversion and considered small tariff changes rather than large ones. But we said nothing about induced changes in the degree of product differentiation (that is, in the number of distinct commodities), and, more prominently yet, imperfect competition and scale economies are foreign to the optimality conditions central to the logic of the propositions. Thus we have set off on some but not all of those departures from standard theory that we have argued for.

We have no wish to construct the awesome structure that would simply add our further suggested innovations to what we have already done in the previous section. Instead we now consider a special model that incorporates our four desired departures in as sharp a fashion as possible: by abstracting from other features. For better or worse, this is also the strategy used in the recent literature on international trade and scale economies, imperfect competition, and product differentiation. The model we use resembles closely some of the models developed in this literature and is basically of the same type as those employed in the two previous chapters. We hence approach our issue from the direction of this recent literature, rather than by modifying the type of models traditionally used in customs unions theory.

We assume there are three countries: the partners (A and B) and the rest of the world, denoted by an asterisk. A and B can be thought of as DCs, the rest of the world as an LDC. A and B produce goods in two common sectors; manufactures and food. There are n and m different variants of manufactures produced in A and B, respectively. They are all (equally) imperfect substitutes to each other from the consumer's point of view, although they are

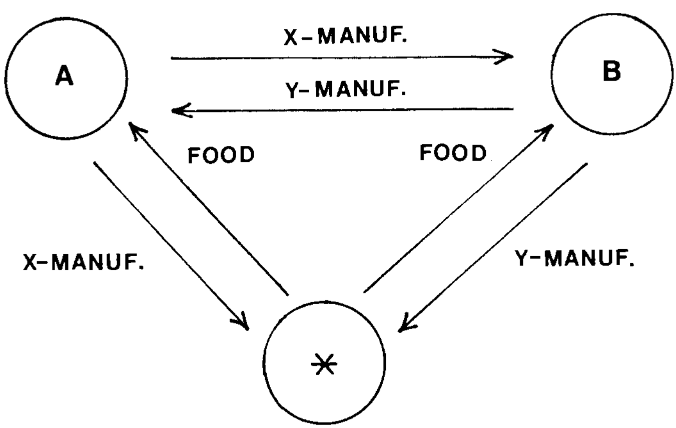

Fig. 1

produced with the same homothetic technology, characterized by internal, increasing returns to scale. Food, on the other hand, is produced in all three countries, under constant returns to scale.

We consider the following special trade pattern. A and B export their respective manufactures to each other and to the rest of the world, and both countries import food from the latter. We assume furthermore that A and B have formed a customs union with their internal tariff τ set at zero, and with a common external tariff t. The rest of the world imposes no trade restrictions.

It is clear that, with such a trade pattern, this customs union does not cause any trade diversion, since the nonmember country exports a different nood than the union countries do. But the union will give rise to trade creation and trade modification.

Following, e.g. Dixit and Stiglitz (1977), we assume the preferences of each country's representative consumer to be given by:

$$U^k = \left(\sum_{i=1}^{n} (X_i^k)^{\beta} + \sum_{j=1}^{m} (Y_j^k)^{\beta} \right)^{\frac{\alpha}{\beta}} (Z^k)^{1-\alpha} \; ; \; k = A, B, * \quad (1)$$

with $0 < \alpha < 1$, and $0 < \beta < 1$, and where X_i^k is the amount of the country A-produced manufacture i that is consumed in country k, Y_j^k is the amount of the country B-produced manufacture j that is consumed in k, and Z^k the consumed quantity of food in k.

Let us now turn to the production side. The production process for manufactures can be thought of as consisting of two stages: first primary factors are used to produce a (non-traded) intermediate input under constant returns to scale, in the amounts M^A and M^B in the two DCs. This input is transformed to final differentiated goods under increasing returns. Since the equilibrium under study is characterized by strong symmetry, in the sense that $X_i^k = X^k$, $\forall i$, and $Y_j^k = Y^k$, $\forall j$, each firm in the manufacturing sector in A employs M^A/n of the intermediate good, and correspondingly M^B/m in B.

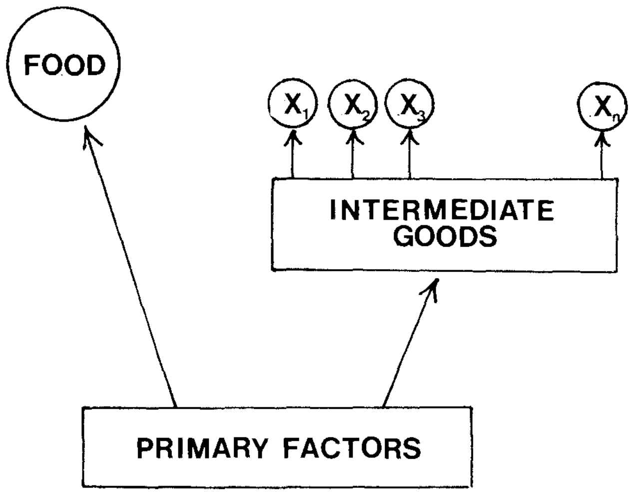

Fig. 2

We assume that entry is free, so the condition for industry equilibrium is that profits are zero:

$$PX = r^A \frac{M^A}{n}, \quad \text{and } QY = r^B \frac{M^B}{m}, \tag{2}$$

where P and Q are the prices received by producers of manufactured products in A and B, respectively, r is the price of the intermediate input, and where $X \equiv X^A + X^B + X^*$, and $Y \equiv Y^A + Y^B + Y^*$. To express the conditions in this form we use the fact that each firm will adjust its quantities in the three markets in such a way that it receives the same price everywhere. The internal increasing returns to scale stem from a fixed cost a, coupled with a constant marginal cost b. We could therefore alternatively simply define

$$M^A \equiv n(a + bX), \quad \text{and } M^B \equiv m(a + bY)$$

and consider r^A, and r^B, as the "factor-price index" in the separable cost fuction.

The conditions for profit maximization in the two manufacturing industries can be stated as:

$$P\beta = r^A b, \text{ and } Q\beta = r^B b. \tag{3}$$

Now, from equations (2) and (3) we immediately get that the total output volume of a firm in the manufacturing sectors is fixed:

$$X = Y = \frac{a}{b} \frac{\beta}{1 - \beta}.$$

This is in particular due to the homothetic technology. It then follows that:

$$n = \frac{1 - \delta}{a} M^A, \quad \text{and} \quad m = \frac{1 - \delta}{a} M^B.$$

The union countries' respective output of food is given by the transformation functions $T(M^A)$ for A, and $S(M^B)$ for B. The domestic produce is sold at the same price as imported food: $P_F(1 + t)$, where P_F denotes the price of food excluding the union's external tariff t. The price of the intermediate good must be the value of the alternative use of the resources devoted to the production of one unit of the good, or:

$$r^A = - P_F(1 + t)\ T'(M^A),$$
$$r^B = - P_F(1 + t)\ S'(M^B). \tag{4}$$

The rest of the world is assumed to use one factor only, with the reward r^*, and with the fixed supply L^*. It takes one unit of this factor to produce one unit of food, and the condition for profit maximization (and industry equilibrium) is therefore simply:

$$P_F = r^*.$$

Two variables that will be of interest in the comparative statics exercises are the internal and external terms-of-trade, $q \equiv \frac{s}{p}$, and $P \equiv \frac{P_F}{p}$, respectively. With the aid of equations (3) and (4) the two variables may be expressed as:

$$q = \frac{S'(M^B)}{T'(M^A)}, \tag{5}$$

$$p = \frac{-1}{b(1 + t)\ T'(M^A)}. \tag{6}$$

Let us finally, before proceeding to the comparative statics exercises, give the clearing conditions for the goods markets (the respective demand functions are given in the appendix):

$$X = \frac{a}{P}\left[\frac{I^A}{n + my_A^{\beta}} + \frac{I^B/(1 + i)}{n + my_B^{\beta}} + \frac{I^*}{n + my_*^{\beta}}\right], \qquad (7)$$

$$Y = \frac{a}{Q}\left[\frac{I^A/(1 + i)}{ny_A^{-\beta} + m} + \frac{I^B}{ny_B^{-\beta} + m} + \frac{I^*}{ny_*^{-\beta} + m}\right], \qquad (8)$$

$$L^* + T(M^A) + S(M^B) = \frac{(1 - a)}{P_F}\left[I^A/(1 + t) + I^B/(1 + t) + I^*\right], \qquad (9)$$

where I^k denotes income of the representative consumer in country k, and where $y_k \equiv Y^k/X^k$.

Comparative statics exercises will be conducted to expose the basic nature of our model. To have a recognizable point of departure and standard of comparison, we shall conduct an exercise similar to that of proposition (3) discussed in section 3. Thus we start with internal free trade ($i = 0$), and with a common positive external tariff ($t > 0$), implying that at the equilibrium point $q = y_k = 1$. We first increase the common external tariff and then increase the common internal tariff. To facilitate matters we make the following assumption: the partners have a transfer scheme such that the relative changes in their incomes I^A and I^B are identical and equal to $\hat{I}$. This reflects our concern with the partners' [illegible] experience, but is not without practical relevance: the common market is an example of a union with an internal transfer scheme.

The rest of this section presents the comparative statics results. Discussions of these results are postponed to the ensuing section 5.

Since we use P_F as the numeraire, we have from equation (9):

$$\hat{I} = -\frac{PnX}{(1-\alpha)(I^A+I^B)}\hat{M}^A - \frac{QmY}{(1-\alpha)(I^A+I^B)}\hat{M}^B + \hat{t} \tag{10}$$

where $\hat{t} \equiv \frac{dt}{1 + t}$.

The numbers of firms in the manufacturing sectors are proportional to the outputs of the two intermediate goods:

$$\hat{n} = \hat{M}^A,$$

$$\hat{m} = \hat{M}^B.$$

The terms-of-trade changes are directly obtained from (5) and (6):

$$\hat{q} = \sigma^B\hat{m} - \sigma^A\hat{n} \ , \tag{11}$$

$$\hat{p} = -\hat{t} - \sigma^A\hat{n} \ , \tag{12}$$

where $\sigma^A \equiv \frac{M^A T''(M^A)}{T'(M^A)}$, and $\sigma^B \equiv \frac{M^B S''(M^B)}{S'(M^B)}$

The changes in trade and consumption patterns are given by:

$$\hat{y}_A = (\sigma^B\hat{m} - \sigma^A\hat{n} + \hat{\tau})/(\beta - 1) \tag{13}$$

$$\hat{y}_B = (\sigma^B\hat{m} - \sigma^A\hat{n} - \hat{\tau})/(\beta - 1) \tag{14}$$

$$\hat{y}_* = (\sigma^B\hat{m} - \sigma^A\hat{n})/(\beta - 1) \tag{15}$$

$$\hat{z}_A = \left(1 + \sigma^A - \frac{m}{n+m}\left(1 - \frac{\beta}{1-\beta}\sigma^A\right)\right)\hat{n} + \frac{m}{n+m}\left(1 - \frac{\beta}{1-\beta}\sigma^B\right)\hat{m} - $$

$$- \frac{m}{n+m}\frac{\beta}{1-\beta}\hat{\tau} \tag{16}$$

$$\hat{z}_B = \left(1 + \sigma^A - \frac{m}{n+m}\left(1 - \frac{\beta}{1-\beta}\sigma^A\right)\right)\hat{n} + \frac{m}{n+m}\left(1 - \frac{\beta}{1-\beta}\sigma^B\right)\hat{m} +$$

$$+ \left(1 + \frac{m}{n+m}\frac{\beta}{1-\beta}\right)\hat{\tau} \tag{17}$$

where $\hat{\tau} \equiv d\tau/(1+\tau)$, and $z_k \equiv Z^k/X^k$ for k = A, B.

The following two expressions are reduced forms of the market equilibrium conditions for X- and Y-goods, respectively, and they will be the starting point of the proceeding analysis:

$$-\theta^*\hat{t} - \left[\theta^B + (\theta^B - \theta^A)\frac{\beta}{(1-\beta)}\frac{m}{(n+m)}\right]\hat{\tau} = \left[\sigma^A + \frac{1}{1-\alpha}\right]\hat{n} +$$

$$+ \frac{m}{n+m}\left[\frac{\beta}{1-\beta}\sigma^A - \frac{1}{1-\alpha}\right](\hat{n}-\hat{m}) +$$

$$+ \frac{m}{(n+m)}\frac{\beta}{(1-\beta)}(\sigma^A - \sigma^B)\hat{m} \tag{18}$$

$$-\theta^*\hat{t} - \left[\theta^A - (\theta^B - \theta^A)\frac{\beta}{(1-\beta)}\frac{n}{(n+m)}\right]\hat{\tau} = \left[\sigma^B + \frac{1}{1-\alpha}\right]\hat{m} -$$

$$- \frac{n}{n+m}\left[\frac{\beta}{1-\beta}\sigma^B - \frac{1}{1-\alpha}\right](\hat{n}-\hat{m}) -$$

$$- \frac{n}{(n+m)}\frac{\beta}{(1-\beta)}(\sigma^A - \sigma^B)\hat{n} \tag{19}$$

where θ^k is the share of world income for country k.

We subsequently make the special assumption that $\sigma^A = \sigma^B \equiv \sigma$, i.e., that the elasticities of the slopes of the transformation curves are the same at the equilibrium point, in the two union countries. This has the merit, aside from increased analytical tractability, of highlighting aspects of particular interest: the importance of the degree of product differentiation and the importance of asymmetries between the member countries in the relative sizes of their manufacturing sectors and also in their relative incomes.

Equations (18) and (19) give the following solution for relative changes in the number of variants produced in the member countries:

$$\hat{n} = -\frac{\theta^*}{\Delta}\hat{t} - \left[\rho^B + (\rho^B - \rho^A)\frac{1}{(1-\alpha)\Gamma}\frac{m}{(n+m)}\right]\frac{1}{\Delta} \tag{20}$$

$$\hat{m} = -\frac{\theta^*}{\Delta}\hat{t} - \left[\rho^A - (\rho^B - \rho^A)\frac{1}{(1-\alpha)\Gamma}\frac{n}{(n+m)}\right]\frac{1}{\Delta} \tag{21}$$

$$\hat{n} - \hat{m} = \frac{(\rho^A - \rho^B)}{\Gamma}\hat{\ } , \tag{22}$$

and

$$\hat{N} = -\frac{\theta^*}{\Delta}\hat{t} - \frac{\Gamma}{\Delta} , \tag{23}$$

where

$$N \equiv n + m,$$

$$\Delta \equiv \sigma + \frac{1}{1-\alpha} > 0,$$

and

$$\Gamma \equiv \rho^A \frac{m}{n+m} + \rho^B \frac{n}{n+m} > 0.$$

The appendix gives some intermediate steps in the derivation of the following two expressions for the relative changes in the indirect utilities V^A and V^*:

$$\hat{V}^A = \alpha\frac{\theta^*}{\Lambda}\Delta\hat{t} + \alpha\left(\frac{\Gamma}{\Lambda}\Delta - \frac{m}{n+m}\right)\hat{\ } \tag{24}$$

$$\hat{V}^* = \alpha\left((\sigma - \frac{1-\beta}{\beta})\frac{\theta^*}{\Lambda} - 1\right)\hat{t} + \alpha(\sigma - \frac{1-\beta}{\beta})\frac{\Gamma}{\Lambda}\hat{\ } , \tag{25}$$

where

$$\Lambda \equiv \sigma - \frac{1-\beta}{\beta} + \frac{1}{(1-\alpha)(\rho^A + \rho^B)} > 0.$$

5. TARIFF CHANGES

Our comparative statics experiments incorporate the two types of changes in union commercial policy that are of interest: changes t in the common external tariff on imports of food, and the marginal imposition of a mutual internal tariff on the exchange of differentiated manufactured goods. Consider each in turn.

5.1 An Increase in the Common External Tariff

Let us first examine the implication of an increase in the already positive common external tariff [illegible] from equations (20) and (21):

$$\hat{n} = \hat{m} < 0.$$

The numbers of varieties of both types of differentiated goods fall as, in each partner country, resources are pulled from manufacturing into the now more highly protected agricultural sector. The contraction of manufacturing is entirely a reduction in the number of product varieties; each produced variety experiences no change in its level of production. With a common external tariff on food, equation (11) implies that $\hat{q} = [illegible](\hat{m}-\hat{n}) = 0$, so that no substitution between product varieties is induced.[4]

Our assumption about the third country ensures that it continues to export d^* units of food. The shift of demand in A and B away from food turns the terms of trade in favor of the union. We have $\hat{p} = -\hat{t} - [illegible]\hat{m}$, and substitution of equation (20) into this expression reveals that the relative price p of food in terms of manufactures does indeed fall, though by less than the tariff increase. The member countries hence have a terms of trade gain to set against the welfare loss from less product variety. But there is, of course, an additional source of gain: the change in income, which is unambiguously positive since the rest of the world is exporting a fixed amount of food. This source is partly reflected by the third term of Λ in equation (24). The direction of change in a member country's welfare is therefore ambiguous. We see that the total effect is more likely to be positive (i) the more sensitive the terms of trade are to given changes in product variety (higher σ), (ii) the less the evaluation of this variety (higher β), and (iii) the larger the share of the rest of the world income (smaller $\Theta^A + \Theta^B$).

The rest of the world has to face both a detrimental terms of trade effect and a reduction of product variety, and is therefore clearly worse off. Algebraically this is immediate from equation (25), noting that $(\sigma\Theta^*/\Delta) - 1 < 0$.

5.2 *A Marginal Internal Barrier*

Now turn to the exercise treated by proposition (3) and suppose that $\hat{t} = 0$ but $\hat{\tau} > 0$. As a starting point, assume that the partners have equal national incomes: $\Theta^A = \Theta^B$. Then again the numbers of varieties produced in the two member countries decline by a common amount, so that $\hat{q} = \sigma(\hat{m} - \hat{n}) = 0$.

Each partner attempts to protect manufacturing, but with resources shifting over to the agricultural sector the result is just the opposite! The reason is as follows. The tariff causes each union member to divert spending from the manufactures of its partner to its own manufactures, and the fact that $\hat{q} = 0$ means that the outside country has no reason to substitute between X goods and Y goods. Thus from equations (13)-(15) we see that $\hat{y}_A = -\hat{\tau}/(1-\beta) < 0$, $\hat{y}_B = \hat{\tau}/(1-\beta) > 0$, and $\hat{y}_* = 0$. As the partners are of equal size, these effects just cancel.

The higher price of the partner's manufactures reduces demand, so that with constant output volumes of individual firms there is downward pressure on the number of firms. This is reinforced by the fact that the proportion $(1-\alpha)$ of the tariff revenue raised in the manufacturing sector is channeled into the agricultural sector. Thus the net effect is to protect food, not manufacturing. From equations (13) and (16) we have that $\hat{y}^A - \hat{z}_A < 0$; that is, a "reversal" of trade modification.

It is instructive to compare section 3's discussion of proposition (3) with behavior in the altered context of the present model. First, recall that in the earlier context gains actually resulted when the reversal of trade modification caused the union to import more from the rest of the world. The structure of the present model allows us to concentrate on what is new by ensuring that such gains shrink to zero: our assumptions about the third country imply that it exports an unchanged quantity of food to the union. Exclusively external trade is neither a substitute for nor complement to internal trade, to use our earlier terminology, and there is also now no trade diversion to "undo". Second, the terms of external trade were assumed fixed in proposition (3), but now the shift of demand toward food turns the terms of trade in the outside world's favor. From equation (12), $\hat{p} = -\sigma\hat{n} > 0$. This could of course be neutralized by an appropriate increase $\hat{t}$ in the common external tariff. But the most significant difference follows from the role of imperfect competition, scale economies, and the endogenous number of product varieties. As in the conventional analysis, zero first-order welfare effects are caused by shifts in union expenditure between existing X goods and Y goods. But in our model the output levels of all product varieties which continue to be produced are unchanged, and instead the number of varieties of manufactures falls. This is deleterious. The total effect on the welfare of country A or B is ambiguous, as can be seen from equation (24). There is now one additional term, compared to the case with the external tariff, which influences the direction of change in welfare: the relative size of the manufacturing sectors. The larger the partner's share of total manufacturing output, the larger the share that will be hit by the tariff, and the more likely it is that the internal tariff will actually reduce welfare, and that a "super-union" will be optimal.

The external country is affected in an ambiguous way. Its terms of trade improve, so that it consumes more manufactures than before along with the same quantity of food. But, on the other hand, those manufactures possess less product diversity than before. In summary, it is possible that all countries lose from the internal tariff, or that all gain, with $Q^A = Q^B$. It is also possible that the member countries gain and the rest of the world loses. But, if the member countries lose, then so must the external country, rendering a "super-union" beneficial to all!

Thus far we have supposed that $Y^A = Y^B$. The role of size disparities between partners is now easily exposed. Equations (20)-(23) show that the internal tariff will still cause a decline in total output of manufactures, but that the smaller country will have to adjust its manufacturing sector more than proportionally: $\hat{n} < \hat{m} < 0$. There is hence a *relocation* of manufacturing from the smaller (in terms of national income) partner to the larger. Intuitively, small-country located firms will have a larger share of production hit by the tariff, inducing more exit to keep profits at zero. Indeed, it might possibly be the case that the larger partner's manufacturing output actually rises, if the difference in national incomes is sufficiently large. This reallocation of production might be one reason why customs unions tend to be formed by similar economies.

5.3 *Intermediate Trade*

There is an alternative interpretation of our model. Assume only two consumption goods, food and a final manufacture, but many differentiated intermediate goods, or components, used to assemble the final manufacture. This interpretation is entirely consistent with our formulation if we let the CES part of the utility function (1) be the assembly function for the final manufacture, X_i^k and Y_j^k represent the inputs of components i and j, and the parameter β reflects the degree of substitutability between different components. The technology is then such that each single producer of an intermediate good faces internal increasing returns to scale available when the production process is geographically concentrated: an example of "national" economies of scale.[5] Assembly of the final manufacture is costless, and exhibits increasing returns to scale in the number of different components that are used. To see this, note that the CES assembly function gives an output of $n^{1/\beta}X$ in the symmetric version, and hence the output increases faster than the input nX. These economies of scale are "international", since the size of the world market for the manufacture determines n, and they also reflect the benefits of an increased division of labor in the world economy; they are assumed to be external to individual assembly firms.

There is, in this interpretation, trade both in intermediate goods, and in the final good food. There is initially a tariff only on the latter good, which is exported by the external country. It is immaterial whether finished manufactures are assembled where consumed, or whether they are all assembled in one partner country and then exported. But in the latter case the interesting possibility of tariffs on final manufactures is not covered by our analysis, the components must be taxed the same whether imported directly or embodied in final manufactures, and exports of the latter must be allowed drawbacks of tariffs paid for imported components. The total impact on each country's welfare of tariff changes is, of course, the same irrespective of model interpretation. What differ are the sources of welfare changes. We saw above that an increase in either the internal or external tariff reduced both n and m unless the partners were too different in terms of their national incomes. Here this means that fewer components will be used in the production of manufactures, leading to a situation with less exploitation of international returns to scale. The main welfare-reducing effect of the tariffs is thus that they are detrimental to the division of labor by diminishing the world market for the final manufacture. But, there is no loss here stemming from reduced exploitation of the traditional, national returns to scale. The two developed countries are trying to protect their home industries, and manage in the sense that a higher proportion of inputs are locally made. But with both countries acting in a similar fashion no over-all gain is ensured.

6. CONCLUDING REMARKS

We live in a multilateral world. Therefore any non-universal change in commercial policy - and they are all in fact non-universal - ought to fall within the domain of the theory of economic intergration. But it has not been so, since that theory usually has been too narrowly conceived. We have argued for four extensions: trade modification, consideration of small changes in policy, scale economies, and product differentiation/imperfect competition.

Adding the first two of these suggested extensions to the conventional framework we derived a basic proposition, suggested by familiar second-best theory, that is central to the characterization of the optimal policy for a subset of countries considering some degree of integration in a tariff-ridden world. We then developed a special, but illustrative, model incorporating all four of our suggested extensions and put it through its paces. Such an exercise is no more than an uncertain first step. But it is in the direction that must be followed, we maintain, if the theory of economic integration is to achieve the relevance to contemporary issued that it long since ought to have possessed.

FOOTNOTES

** The authors received useful comments from Elhaman Helpman and from participants at a seminar at the Institute for International Economic Studies, University of Stockholm. Henrik Horn gratefully acknowledges financial support from Humanistisk-Samhällsvetenskapliga Forskningsrådet, Sweden.

1. See Meade (1955, 1956) and Ethier (1983, chapter 12).

2. For recent examples see Wonnacott and Wonnacott (1981), and Lloyd (1982).

3. The most notable exception is no doubt James Meade (1955, 1956), who was criticised in this regard. For example Lipsey (1967,pp. 271) refers to this "very serious, possibly crippling, limitation". The subsequent literature has largely avoided marginal analysis. See, however, Berglas (1979), Reizman (1979), McMillan and McCarr (1981), and Lloyd (1982).

4. In the more general case $\sigma^A \neq \sigma^B$ we would have $\sigma^A \hat{n} = \sigma^B \hat{m}$ so that manufacturing contracts relatively more in the country with the greater ease of substitution of food for the nontraded manufacturing intermediate good; the structure of such economies remains irrelevant. Also we would still have $0 = \hat{q} = \sigma^B \hat{m} - \sigma^A \hat{n}$.

5. See Ethier (1979, 1982).

APPENDIX 1*

This section supplies an explicit proof of proposition (3) and discusses the proposition further. Assume three countries: two partners (A and B) and the rest of the world (distinguished by an asterisk). There is a total of n traded goods, divided into the twelve groups described in Table 1. Groups 1 to 5 contain goods imported by country A, exportables are collected in groups 6-10, whereas country A doesn't trade goods in groups 11 and 12.

Group	*Commodity Description*
1	B exports only to A
2	B exports to A and to the rest of the world
3	the rest of the world exports only to A
4	the rest of the world exports to A and to B
5	the rest of the world and B both export to A
6	A exports only to B
7	the rest of the world and A both export to B
8	A exports only to the rest of the world
9	A and B both export to the rest of the world
10	A exports to both B and the rest of the world
11	B exports only to the rest of the world
12	the rest of the world exports only to B

Table 1. Description of the commodity groups

* Appendix 1 is written by Wilfred Ethier.

For each group i, Q_i denotes the n-vector with a zero in each component corresponding to a good not in group, i, and with the country A domestic price of the respective good in each component that does so correspond. Thus $Q = \sum_{i=1}^{12} Q_i$ is the vector of country A domestic prices. Vectors P_i apply analogously to the partner country, so $P = \sum_{i=1}^{12} P_i$ denotes the vector of country-B domestic prices.

In like fashion, P_i^* denotes the vector of prices actually paid to the rest of the world, or received from the rest of the world, for goods in group i (with zeroes in components corresponding to other goods). Let $R^A = P_1 + P_2^* + P_3^* + P_4^* + P_5^* + Q_6 + Q_7 + P_8^* + P_9^* + P_{10}^*$. Then R_A is the vector of prices at which country A actually transacts with foreigners, *except* possibly for those goods in group 5 that are imported from country B. In like manner, $R^B = P_1 + P_2^* + P_4^* + P_5 + Q_6 + P_7^* + P_9^* + P_{10}^* + P_{11}^* + P_{12}^*$ denotes B's actual international transaction prices, *except* possibly for goods in group 7 imported from A.

T_i^A, for i = 1,..., 4, is the n x n diagonal matrix with ad-valorem tariff rates of country A on goods in group i as the appropriate diagonal elements, and zeroes elsewhere. T_{5B}^A and T_{5*}^A likewise denote tariff matrices of country A on goods in group 5 imported from B and from the rest of the world respectively. Country B tariff matrices T_4^B, T_6^B, T_{10}^B, T_{12}^B, T_{7A}^B, and T_{7*}^B are defined analogously. Let $T^A = T_1^A + T_2^A + T_3^A + T_4^A + T_{5*}^A$ and $T^B = T_4^B + T_6^B + T_{7*}^B + T_{10}^B + T_{12}^B$. Then

$$Q = (I + T^A)\ R^A,$$

$$P = (I + T^B)\ R^B.$$

Note also that $(I + T^A_{5B})\ P_5 = (I + T^A_{5*})\ P^*_5$ and $(I + T^B_{7A})\ Q_7 = = (I + T^B_{7*})\ P^*_7$.

M^A_i is the n-vector of country A net imports of goods in group i (with zeroes elsewhere), and $M^A = \sum_{i=1}^{10} M^A_i$. Also $M^A_5 = M^A_{5B} + M^A_{5*}$ and $M^A_{10} = M^A_{10B} + M^A_{10*}$ where the two vectors on the right in each case record trade with country B and the rest of the world respectively. Analogously for M^B_i, for $M^B = M^B_1 + M^B_2 + M^B_4 + M^B_5 + M^B_6 + M^B_7 + M^B_9 - M^B_{10} - M^B_{11} + M^B_{12}$, and for both $M^B_2 = M^B_{2A} + M^B_{2*}$ and $M^B_7 = M^B_{7A} + M^B_{7*}$. (Note that $M^A_1 = - M^B_1$, $M^A_2 = - M^B_{2A}$, $M^A_{5B} = - M^B_5$, $- M^A_6 = M^B_6$, $- M^A_7 = M^B_{7A}$ and $- M^A_{10B} = M^B_{10}$).

In country A, total expenditure - denoted by the national expenditure function $e^A(Q, u_A)$, where u_A denotes the vector of utilities of A residents - must equal income, composed of tariff revenue plus the value of national production, the latter given by the national product function $y^A(Q)$. Thus

$$e^A(Q, u_A) = y^A(Q) + T^A\ R^A\ [M^A - M^A_{5B}] + T^A_{5B}\ P^A_5\ M^A_{5B}. \tag{A1}$$

Analogously for the partner country:

$$e^B(P, u_B) = y^B(P) + T^B\ R^B\ [M^B - M^B_{7A}] + T^B_{7A}\ Q_7\ M^B_{7A}. \tag{A2}$$

A Basic Expression

Differentiating (A1), noting that $\partial e^A/\partial Q - \partial y^A/\partial Q = M^A$ and that $dQ = dR^A + d(T^A\ R^A)$ gives:

$$\frac{\partial e^A}{\partial u_A} du_A = - M^A \, dR^A - M^A_{5B} \, [d(T^A \, R^A) - d(T^A_{5B} \, P_5)] + T^A \, R^A \, d\bar{M}^A + $$
$$+ T^A_{5B} \, P_5 \, dM^A_{5B}, \qquad (A3)$$

where $\bar{M}^A \equiv M^A - M^A_{5B}$. In a like manner, we obtain for the partner country

$$\frac{\partial e^B}{\partial u_B} du_B = - M^B \, dR^B - M^B_{7A} \, [d(T^B \, R^B) - d(T^B_{7A} \, Q_7)] + T^B \, R^B \, d\bar{M}^B + $$
$$+ T^B_{7A} \, Q_7 \, dM^B_{7A}. \qquad (A4)$$

Equations (A3) and (A4) supply measures of the welfare effects on the partners of changes in commercial policy. The sources of these effects, spelled out on the right-hand sides of the equations, are analogous to the sources of welfare effects commonly discussed in conventional tariff theory. (For such a discussion, see section A.6 of Appendix I in Ethier (1983).).

We measure the change in the joint welfare of the partner countries by $dW = \frac{\partial e^A}{\partial u_A} du_A + \frac{\partial e^B}{\partial u_B} du_B$. Remember that we suppose constancy in the prices at which A and B transact with the rest of the world, implying:

$$M^A \, dR^A + M^B \, dR^B = - M^A_{5B} \, dP_5 - M^B_{7A} \, dQ_7.$$

With goods in group 5 (7) being bought from the rest of the world at prices assumed to be unchanged, $d(T^A \, R^A)$ $(d(T^B \, R^B))$ will have zero elements corresponding to positive entries in M^A_{5B} (M^B_{7A}), and therefore:

$$dW = M^A_{5B} \, [dP_5 + d(T^A_{5B} \, P_5)] + T^A_{5B} \, P_5(dM^A - d\bar{M}^A) + T^A \, R^A \, d\bar{M}^A + $$
$$+ M^B_{7A} \, [dQ_7 + d(T^B_{7A} \, Q_7)] + T^B_{7A} \, Q_7(dM^B - d\bar{M}^B) + T^B \, R^B \, d\bar{M}^B, $$
$$(A5)$$

where $T^A_{5B}\ P_5 (T^B_{7A}\ Q_7)$ has zeroes corresponding to non-zero elements of $d\bar{M}^A$ $(d\bar{M}^B)$. Furthermore,

$$dP_5 + d(T^A_{5B}\ P_5) = d[(I + T^A_{5B})\ P_5] = d[(I + T^A_{5*})\ P^*_5] \doteq P^*_5\ dT^A_{5*},$$

$$dQ_7 + d(T^A_{7A}\ Q_7) = d[(I + T^B_{7A})\ Q_7] = d[(I + T^B_{7*})\ P^*_7] = P^*_7\ dT^B_{7*}.$$

Since also M^A_{5B} (M^B_{7A}) has zero elements corresponding to non-zero elements in $P^*_5\ dT^A_{5*}$ $(P^*_7\ dT^B_{7*})$, our expression for the effect of tariff changes on the partners' joint welfare simplifies to

$$dW = (T^A_{5B}\ P_5)\ dM^A + (T^B_{7A}\ Q_7)\ dM^B + (T^A\ R^A)\ d\bar{M}^A + (T^B\ R^B)\ d\bar{M}^B. \tag{A6}$$

This is our basic result. We now apply it to two special cases in order to derive proposition (3).

The Proposition

Suppose that A and B initially have a full free-trade area and impose marginal internal tariffs. Then $T^A_1 = T^A_2 = T^A_{5B} = T^B_6 = T^B_{7A} = T^B_{10} = 0$ and (A6) reduces to

$$dW = (T^A\ R^A)\ d\bar{M}^A + (T^B\ R^B)\ d\bar{M}^B. \tag{A7}$$

Thus a marginal retreat from internal free trade will raise the partners' joint welfare if and only if it raises total tariff revenue collected on trade with the rest of the world, as stated above in proposition (3). Note in particular two aspects of this result.

First, the degree of trade diversion - reflected in the *composition* of dM^A_5 between dM^A_{5B} and dM^A_{5*} and in the composition of dM_7^B between dM_{7A}^B and dM_{7*}^B - has no influence at all on dW. This is to be expected in the light of the previous section's informal discussion.

Second, trade modification - reflected in the magnitudes of dM_3^A, dM_4^A, dM_4^B, and dM_{12}^B - will add to or subtract from dW according as the goods involved are in this sense complementary to or substitutes for intra-union trade. This, again, is not surprising in view of the earlier discussion. Indeed, if the goods which countries A and B trade exclusively with the rest of the world are sufficiently strong substitutes for the goods they exchange with each other, they could have an incentive to form an "anti-union", that is, to add a marginal discriminatory tariff on each others' goods. The two would have reason to seek separate preferential arrangements with the rest of the world, rather than with each.

It is instructive to decompose the right-head side of (A7) as follows:

$$dW = (T_{5*}^A P_5^* dM_{5*}^A + T_{7*}^B P_7^* dM_{7*}^B) + (T_3^A P_3^* dM_3^A + T_4^A P_4^* dM_4^A + + T_4^B P_4^* dM_4^B + T_{12}^B P_{12}^* dM_{12}^B) \qquad (A8)$$

Note that an "undoing" of trade diversion - indicated by positive terms within the first set of parentheses on the right-hand side of (A8) - must increase W regardless of how great the accompanying reduction in intra-union trade in the same goods happens to be. This is a consequence of the initial free internal trade, as discussed in the previous section.

The influence of trade modification is summarized by the second parenthesized term on the right-hand side of (A8). Joint partner welfare is thereby raised or lowered according as this term is positive or negative, that is, according as the goods traded exclusively with the rest of the world are in this sense on balance substitutes for or complementary to intra-partner trade.

Trade diversion and trade modification is central in the exercise. The presumptive role of trade diversion is clear; it should cause a marginal retreat from a free-trade area to raise joint partner welfare. But the role of trade modification depends crucially upon whether the goods traded exclusively with the rest of the world are on balance substitutes for or complements to intra-union trade, in the sense that discrimination in favor of the latter reduces or increases the volume of the former, with the change in this volume measured by the net change in partner tariff revenue it would generate at initial prices and tariffs.

A free trade area would not constitute an optimal policy - for A and B jointly, given tariffs on goods from the rest of the world - without a very special balancing of the various influences discussed above. If exclusively external trade is on balance neither strongly substitutable for nor complementary to internal trade, optimal policy for the partners will presumably entail tariff preferences that fall short of a full free trade area, so that customs unions should be unstable.

Significant net *substitutability* both reduces the gain from initial preferences and increases the gain to a retreat from a full union. Thus one expects it to cause optimal policy to feature a lower degree of preference. Sufficient substitutability should make optimal an "anti union" with the members actually discriminating against each other. Significant net *complementarity*, finally, increasing the gain from initial preferences and reducing that from a retreat, presumably cause the optimal policy to more closely approximate free internal trade. Large enough complementarity could cause the optimal policy to be a "super union" with internal trade actually stimulated by subsidies financed by tariffs on external trade. Note that the European Community's Common Agricultural Policy - which does not naturally come to mind in a discussion of optimality - does in fact possess some of the characteristics of such a case.

APPENDIX 2

The Demand Functions

$$PX^A = \frac{\alpha\ I^A}{n + m\ y_A^{\beta}} \qquad \text{where } y_A \equiv \frac{Y^A}{X^A}$$

$$Q(1 + \tau)Y^A = \frac{\alpha\ I^A}{ny_A^{-\beta} + m}$$

$$P_F(1 + t)Z^A = (1 - \alpha)I^A$$

$$P(1 + \tau)X^B = \frac{\alpha\ I^B}{n + my_B^{\beta}} \qquad \text{where } y_B \equiv \frac{Y^B}{X^B}$$

$$QY^B = \frac{\alpha\ I^B}{ny_B^{-\beta} + m}$$

$$P_F(1 + t)Z^B = (1 - \alpha)I^B$$

$$PX^* = \frac{\alpha\ I^*}{n + my_*^{\beta}} \qquad \text{where } y_* \equiv \frac{Y^*}{X^*}$$

$$QY^* = \frac{\alpha\ I^*}{ny_*^{-\beta} + m}$$

$$P_F\ Z^* = (1 - \alpha)I^*$$

The Relative Change in the Indirect Utilities

The indirect utility function for country A is:

$$V^A = \gamma(R^A)^{-\alpha}\ (1 + t)^{-(1-\alpha)}\ I^A,$$

where $R^A \equiv \left(nP^{\rho} + m(1 + \tau)^{\rho}Q^{\rho}\right)^{\frac{1}{\rho}} = p^{-1}\left(n + m(1 + \tau)^{\rho}q^{\rho}\right)^{\frac{1}{\rho}},$

and $\rho \equiv \frac{\beta}{\beta - 1}$,

and for the third country

$$V^* = \gamma\ (R^*)^{-\alpha}\ I^*,$$

where

$$R^* = p^{-1}(n + mq^{\rho})^{\frac{1}{\rho}}.$$

γ is a constant, and R^k is a price index defined over the differentiated goods. These expressions become, when differentiated logarithmically:

$$\hat{V}^A = - \alpha\ \hat{R}^A - (1 - \alpha)\hat{t} + \hat{I},$$

and

$$\hat{V}^* = - \alpha\ \hat{R}^*\ ,$$

where we use that $\hat{I}^A = \hat{I}^B \equiv \hat{I}$, and that $\hat{I}^* = 0$.

Concentrate for a moment on the change in the price index R^A. The relative change can be split up into two components:

$$\hat{R}^A = \left[\frac{m}{n + m}\ (\hat{q} + \hat{\tau}) - \hat{p}\right] - \frac{1 - \beta}{\beta}\ \hat{N}. \tag{A9}$$

The second term is the effect on the index of a change in the number of varieties available to the consumer, at given relative prices. This term is of course not present in the traditional analysis of customs unions, but will prove to be of considerable importance here.

Returning to the change in utility we get by substitution:

$$\hat{V}^A = - \left[\hat{t} + \alpha\frac{m}{n + m}\ \hat{\tau} + \alpha\sigma\hat{N}\right] + \hat{I} + \alpha\ \frac{1 - \beta}{\beta}\ \hat{N}. \tag{A10}$$

By substituting equations (10) and (23) into (A10) we arrive at expressions (24) and (25) in the text.

REFERENCES

Berglas, E., 1979, "Preferential trading theory: The n commodity case", *Journal of Political Economy* 87, pp 315-331.

Dixit, A.K., and J. Stiglitz, 1977, "Monopolistic competition and optimum product diversity", *American Economic Review* 67, pp 297-308.

Ethier, W., 1979, "Internationally decreasing costs and world trade", *Journal of International Economics* 9, pp 1-24.

Ethier, W.J., 1982, "National and international returns to scale in the modern theory of international trade", *American Economic Review* 72, pp 389-405.

Ethier, W.J., 1983, *Modern International Economics*, W.W. Norton, New York.

Kemp, M.C., 1964, *The Pure Theory of International Trade*, Prentice-Hall, Englewood Cliffs.

Kemp, M.C., and H.Y. Wan, 1976, "An elementary proposition concerning the formation of customs unions", *Journal of International Economics* 6, pp 95-98.

Kravis, I.B. and R.E. Lipsey, 1971, *Price competitiveness in world trade*, National Bureau of Economic Research, New York.

Kravis, I.B. and R.E. Lipsey, 1978, "Price behavior in the light of balance of payments theories", *Journal of International Economics* 8, pp 193-246.

Lipsey, R.E. 1968, "The theory of customs unions: A general survey", in R.E. Caves and H.G. Johanson (eds.), *Readings in International Economics*, Richard D. Irwin, Homewood.

Lloyd, P., 1982, "3 x 3 theory of customs unions", *Journal of International Economics* 12, pp 41-63.

McMillan, J. and E. McCann, 1981, "Welfare effects in customs unions", *Ecoonomic Journal* 91, pp 697-703.

Meade, J.E., 1955, *Trade and welfare*, Oxford University Press, London.

Meade, J.E., 1956, *The Theory of Customs Unions*, North Holland, Amsterdam.

Ohyama, M., 1972, "Trade and Welfare in General Equilibrium", *Keio Economic Studies*, 9, pp 37-73.

Reizman, R., 1979, "A 3 x 3 model of customs unions", *Journal of International Economics* 9, pp 341-354.

Vanek, J., 1965, *General Equilibrium of International Discrimination*, Harvard University Press, Cambridge.

Viner, J., 1950, *The Customs Union Issue*, Carnegie Endowment for International Press, New York.

Wonnacot, P., and R. Wonnacot, 1981, "Is unilateral tariff reduction preferable to a customs union? The curious case of the missing foreign tariffs", *American Economic Review* 71, pp 704-714.

CHAPTER V

TRADE UNION DETERMINED WAGES, UNEMPLOYMENT, AND THE SIZE OF THE PUBLIC SECTOR

CHAPTER V TRADE UNION DETERMINED WAGES, UNEMPLOYMENT, AND THE SIZE OF THE PUBLIC SECTOR*

1. INTRODUCTION

Most smaller developed market economies are characterized by a large public sector, and by the fact that the seller's side of the labor market is controlled by a small number of trade unions, or confederations of trade unions, organizing a large share of workers. Intuitively, one would expect the size and the growth of the public sector to have exerted influence on these unions' wage claims through at least two channels. First, the financing of this growth has required continuously increasing taxes. In many countries industrial workers' average income tax rates amount to 30 to 40 %, and so do pay-roll taxes. At the margin these taxes are even more striking. For instance, Lindbeck (1983a) reports that the marginal wedge between the return to a household and the wage cost for a firm often is around 300 % and occasionally more than 700 % in Sweden. Taxes have risen to the extent that it is sometimes asserted that government tax policies are equally or more important for workers' real income than is the outcome of wage settlements.[1] Second, the public sector's expansion has also had the effect of diminishing the share of the work force that is employed by the private sector. It has hence reduced the number of employees working in firms that exhibit e.g. cyclical variations in demand for labor. This effect is potentially of considerable importance, since the public sector's share of total employment is around 20 % in many countries, and is approaching and exceeding 30 % in some. But even though a close relationship between union wage claims and the size of the public sector is frequently emphasized in the more policy-oriented debate, very little attention has been directed by theorists to the possible nature of such linkages.[2]

The purpose of this chapter is to theoretically analyze some aspects of the relationship between the size of the public sector, trade union determined wages, and unemployment. Three issues are investigated, in particular. The first question considered is the relationship between the absolute size of the public sector and the economy's unemployment rate.

The second issue is inspired by the demands for a reduced public sector that are increasingly often put forward in developed countries as remedies for large and accumulating government budget deficits. These proposals are frequently objected to on the grounds that such reductions have to be so substantial that they will lead to serious unemployment problems, if they are to have any noticeable effect on the budget deficit. It appears as if there is a fear that this unemployment is not only a temporary, search phenomenon, but that it is of a more persistent or "structural" type. Can the fear for longer run unemployment be formally substantiated?

The third question considered is also frequently discussed among at least more policy-oriented European economists. It concerns possible longer run allocational effects of short run stabilization policies of a type pursued in the 1960s and 70s, in particular. These policies take the form of more or less unconditional interventions by the government into the labor market; unconditional in the sense that they are undertaken regardless of the source of the unemployment problem. Governments have hence attempted to stabilize employment by counter-cyclical hirings and layoffs. But it is sometimes asserted that governments were more successful in hiring workers in economic slumps, than in laying them off in times of higher labor demand. It is even sometimes claimed that the very substantial expansion of the public sector in many European countries, can largely be explained by this asymmetry.[3] The third issue considered is possible allocational effects of an employment stabilization policy of this type.

The wage formation process can evidently be viewed from different perspectives, each focussing on a particular aspect. The one emphasized here is the potential *divergence in interests* between workers who already have established themselves in the private sector labor market, and those who may enter this market. It turns out that such a divergence is potentially very important.

The chapter's basic idea is that workers who already have a job in the private sector have a higher chance of being employed in the (near) future, than do workers who are about to enter this labor market. One can advance at least two reasons why this should be the case. First, there is usually some type of seniority system which determines the order in which workers are laid off and rehired. Second, the possibility of acquiring work experience gives the already employed an advantage over entrants, in particular when wages for entrants are higher than their value marginal product, as is often argued to be the case.

The fact that workers differ in their probability of being employed implies that they will have different most-preferred wages, when faced with a trade-off between private consumption and the probability of employment. If workers perceive the world as classical, i.e. that there is a negative relation between the product wage and the level of employment, those who already hold jobs will favor higher wages than will entrants.

Two features of trade union organizations then become important. The first is that only trade union members can influence the union's wage policy, and the second is that to become a member it is usually required that one already has a job in the sector.[4] The combination of these factors results in a system where the private sector's uniform wage is not determined in the interest of *all* those who want a private sector job. Instead, only those who are already established in the private sector have an opportunity of influencing the wage, and the wage will therefore be higher than it would otherwise be.

The model is presented in section 2. The object of analysis is a closed economy, consisting of a private and a public sector. To highlight union influences on the economy, the union is assumed to unilaterally set the wage. The private sector experiences varying levels of unemployment, due to a random component in the private sector's demand for labor. To capture the fact that the probability of future employment differs between workers who are already employed and workers who are new in the labor market, we adopt a slightly modified version of a trade union model developed by Grossman (1983a, b), referred to as the "seniority model". It is assumed that union members differ in seniority, and that possible layoffs or rehirings are determined according to this ranking. The uniform wage is determined through majority voting with direct democracy. At the time of voting the demand for labor is uncertain, but the uncertainty is resolved immediately after the wage is set. Preferences of union members are single-peaked, and hence the outcome of the voting is the wage that is most preferred by the member who is the median in terms of the seniority ranking. Possible unemployment is hence "classical", resulting from too high real wages.

The rest of the chapter is structured as follows. The trade union's reaction function is derived in section 3. The ensuing section considers how the size of the public sector influences "long run" wages and unemployment in the private sector. Section 5 studies some dynamics of wages and unemployment that result from a once-and-for-all reduction in the size of the public sector. In section 6 structural consequences of a government employment stabilization policy are analyzed. Section 7, finally, contains a tentative discussion of some policies which aim at reducing the economy's unemployment, and also some concluding comments.

2. THE MODEL

Consider an economy comprising a public and a private sector. Both sectors use the only variable factor, labor, which is homogenous and perfectly mobile between sectors. The *public sector* influences the economy in two ways. First, it employs G workers, where the level of G is determined by an exogenous political decision. Second, the government levies a proportional tax τ on each employed worker. The tax may vary positively with the size of the public sector, or it may be constant[5,6]:

$$\tau = T(G), \qquad 0 \leq \tau < 1, \qquad T_G \geq 0. \qquad (1)$$

Firms are perfectly competitive in the private sector. The industry's demand for labor N depends on the real (product) wage w and on a random productivity parameter α. We assume for convenience that α is uniformly distributed in the interval $[1 - \delta, 1 + \delta]$, and that the deterministic part of the demand function is linear in the wage[7,8]. Formally,

$$N = N(w, \alpha) = \alpha\nu(w), \quad \nu_w < 0, \quad \nu_{ww} = 0.$$

The actor of particular interest to us is the *private sector trade union*. There might be a public sector trade union as well, but since we are not concerned with the interaction between trade unions, and since the size of public employment is exogenously determined, we focus on the private sector union only. The latter organizes all privately employed workers. There is in our formulation no explicit incentive for a worker to join the union, but we assume he has to become a member of the union, when offered a private sector job, to be permitted to work in the sector. We assume also, for simplicity, that a worker who is a union member remains in the union unless he becomes employed by the government, even if he is temporarily laid off in the private sector. A worker who have lost his public sector job does not become a trade union member until he has obtained a job in the private sector - a private sector job offer is a prerequisite for union membership.

The basic assumptions of the seniority model are that the wage is unilaterally set by the union before the productivity level (and therefore labor demand) is known, and that seniority determines who will be laid off or rehired[9]. Hence, workers are not the same even *ex ante* the wage decision, since they are assumed to differ in their seniority ranking. A worker's ranking is given by an index i which lies in the interval [0 , M], where M is the size of the union's membership at the time of voting and where i = 0 is the most senior worker. Each worker has a unique most-preferred wage depending on his seniority ranking, with more senior workers preferring higher wages. The union's wage demand is determined according to a majority voting rule. It can be shown that there exists under these conditions a majority voting equilibrium which maximizes the expected utility of the median voter (i.e. the worker with seniority ranking M/2) [10].

To keep matters as simple as possible, we assume that workers derive utility only from private consumption. Let V(c) depict a worker's utility function, which is assumed to be increasing but at a decreasing rate in the consumption level c. There are no savings, so for the employed worker c is equal to the after-tax real income $(1 - \tau)w$, and for the unemployed c equals the fixed level b.

Worker i will be employed if the realized demand for labor is larger than or equal to the number of workers more senior than him. The probability for this to occur is [11]:

$$\pi^i = \Pr(N(w , \alpha) \geq i) =$$

$$= \min\left[\frac{1}{2\delta}\left(1 + \delta - \frac{i}{\nu(w)}\right), 1\right] . \qquad (3)$$

The outcome of the majority vote is therefore the wage which solves:

$$\max_w \; U^{M/2} = \pi^{M/2} V((1 - \tau)w) + (1 - \pi^{M/2}) V(b)$$

$$\text{s.t. } \pi^{M/2} = \frac{1}{2\delta}\left[1 + \delta - \frac{M/2}{\nu(w)}\right]. \qquad (4)$$

We assume that the worker treats the government's decision variables as exogenously given and that he disregards the effects of induced future changes in the size of the membership, when deciding on his optimal wage. Assuming an interior solution, the optimization problem for the median voter gives rise to the following first-order condition[12]:

$$\pi^{M/2} \frac{d}{dw} [V((1 - \tau)w)] + [V((1 - \tau)w) - V(b)] \frac{d\pi^{M/2}}{dw} = 0,$$

or, explicitly expressed,

$$\Lambda(w, \tau, M) \equiv (1 - \tau)\left(1 + \delta - \frac{M/2}{\nu(w)}\right) V_c((1 - \tau)w) + \\ + \frac{M/2}{\nu(w)^2} \nu_w(w)[V((1 - \tau)w) - V(b)] = 0. \qquad (5)$$

Consider, say, an increase in the wage with one unit. This will have two effects for the worker. First, he will for a constant probability of employment, increase his consumption when employed with $1 - \tau$, which obviously raises his utility. This is the first term in equation (5). Second, his probability of employment is reduced, since for each realized productivity level the demand for labor is reduced by $\alpha\nu_w$. The utility loss from this depends on the difference in consumption levels when employed and when unemployed, as given by the second term.

3. THE TRADE UNION'S REACTION FUNCTION

As mentioned in the introduction, the government influences the union's choice of wage in two ways: it levies an income tax on workers, and it determines the size of the public sector, which may affect the union's membership. Our next task is to determine the nature of these relationships.

Equation (5) gives the median voter's voted wage w as an implicit function of, among other variables, the tax rate and the size of the membership. We express these relationships compactly in the form of a reaction function, $w = \phi(\tau , M)$.

Let us start by considering how the tax affects the voted wage. Differentiation of the first-order condition (5) gives[13]:

$$\frac{\partial w}{\partial \tau} \equiv \phi_\tau(\tau , M) = - \frac{\Lambda_\tau}{\Lambda_w} , \qquad (6)$$

where

$$\Lambda_\tau = - \left(1 + \delta - \frac{M/2}{\nu}\right)\left(1 + \frac{(1-\tau)w\ V_{cc}}{V_c}\right) V_c - \frac{M/2}{\nu^2}\, w\, \nu_w\, V_c .$$

Λ_w is negative, fullfilling the second-order condition, but Λ_τ can in principle be of either sign. The latter is unambiguously positive if the absolute value of the coefficient of relative risk aversion $(1 - \tau)w\ V_{cc}/V_c$, is larger than unity (cf Oswald (1982)). There are three effects involved; two of which tend to give a positive relation between the tax and the voted wage, and one which works in the opposite direction. The negative effect stems from the term $- (1 + \delta - \frac{M/2}{\nu})\ V_c$. It captures the "proportional" as opposed to "lump-sum" nature of the tax: an increased proportional tax implies that a wage increase is less valuable, since less of the increase in income is kept with the higher tax. A lower wage will therefore be preferred, *ceteris paribus*, since the price payed for the last wage unit in the terms of a lower probability of employment is the same as before the tax increase. This term is not present if the tax is lump-sum, and there is therefore in this case an unambiguously positive relation between the voted wage and the tax.

The first positive effect results from that the increase in the tax increases the marginal utility of consumption when employed. This is the term including the coefficient of relative risk aversion. The third term, finally, captures that the difference between utility levels when employed and when unemployed is diminished. Both these terms hence tend to increase the voted wage.

We assume that the tendency to lower the wage is dominated by the tax-push effects, working in the opposite direction, i.e. that $\phi_\tau > 0$. It seems that the empirically more interesting case is the one when the trade union seeks to compensate itself, at least partially, for tax increases. This view seems to be well in accordance with conventional wisdom among economists; see e.g. Lindbeck (1983b)[14].

The second relation of particular interest is that between the size of the membership and the majority chosen wage. We derive

$$\frac{\partial w}{\partial M} = \phi_M = - \frac{\Lambda_M}{\Lambda_w} < 0, \tag{7}$$

with

$$\Lambda_M = - \frac{(1 - \tau)}{2\nu} V_c + \frac{\nu_w}{2\nu^2} \bar{V} < 0,$$

and

$$\bar{V} \equiv V\Big((1 - \tau)w\Big) - V(b);$$

that is, if additional members join the union, it will select a lower wage; a result earlier derived by Grossman (1983a). Intuitively, the larger the union, the further from the top is the median voter, and the less his chance of becoming employed at a given wage.

It seems that a negative relation between the size of the membership and the wage should be expected for any type of trade union facing a downward sloping demand curve for labor, at least for infra-marginal changes in the size of the membership. If a union is concerned about the employment *rate* of its membership, then the entry of new members forces the union to set a lower wage, which provides absolutely more jobs.

Let us make a final assumption before considering the first experiment. We assume that the voted wage is such that the least senior worker has a positive probability of being employed. Union membership would otherwise imply no advantage at all to this worker, and he would therefore probably leave the union. This assumption implies that labor demand in the state with the highest productivity level must be at least as large as the size of the union. We therefore rule out the set of combinations of the wage, the membership, and the maximum productivity level, fullfilling the inequality $N(w, 1 + \delta) < M$. The boundary of this set is given by $N(w, 1 + \delta) = M$, which we denote the "Positive Probability" function (PP), since the marginal member's probability of employment is just positive for combinations of w and M fullfilling the equation.

The PP-function might be more or less steep than the reaction function. To distinguish the two possibilities we denote the union portrayed in Fig. 1 "stable", whereas Fig. 2 depicts an "unstable" union. (The choice of this terminology is made clear below.) The assumption that the marginal member has a strictly positive probability of employment hence implies that we restrict the analysis of the stable union to cases to the left of the intersection of the two schedules in Fig. 1., i.e. two cases where the union is smaller than $\hat{M}$. Otherwise some members wouldn't get a job in any state of nature.

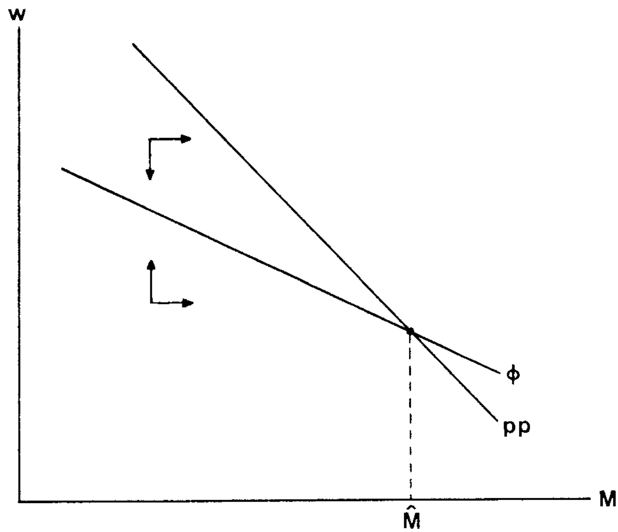

Figure 1

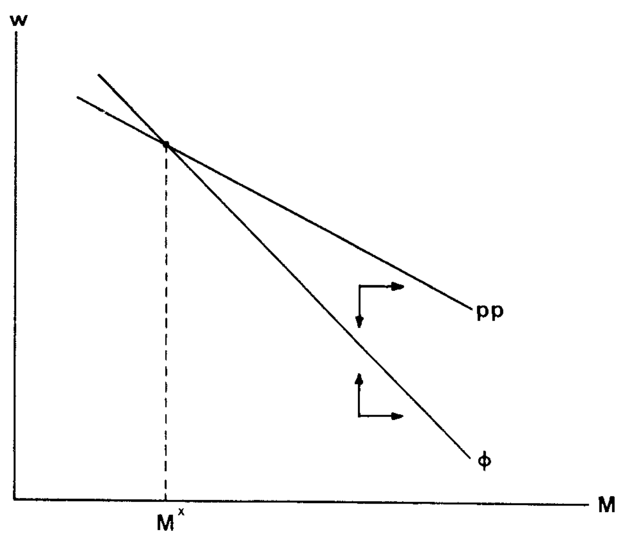

Figure 2

Correspondingly, if the union is unstable, we concentrate on situations where the union is larger than M^x in Fig. 2.

4. THE SIZE OF THE PUBLIC SECTOR, WAGES AND UNEMPLOYMENT

In this section we will consider the relationship between the size of the public sector, wages and unemployment. Even though the results follow fairly directly from the previous section, they nevertheless deserve to be pointed out. They serve also as a background for the next section's study of dynamic properties of the model.

We will subsequently refer to a situation where all workers are either union members or government employees, and so remain period after period, as a "stationary equilibrium". In such an equilibrium the voted wage hence also remains the same, whereas private sector employment varies with the random productivity level. Let us start by comparing two economies with the only exogenous difference that in economy "s", the size of the public sector is larger than in economy "b"[15]. Both economies are in a stationary equilibrium. Letting L denote the total constant work force:

$$M^s = L - G^s < L - G^b = M^b; \qquad (8)$$

that is, the union membership is smaller in s than in b, since all non-government employees are members of the trade union. Furthermore, it is clear that the tax rate is higher in s than in b, unless $T_G = 0$:

$$\tau^s = T(G^s) \geq T(G^b) = \tau^b \quad \text{for} \quad T_G \geq 0. \qquad (9)$$

It is then immediate from the previous section that *the wage is higher in economy s than in economy b:*

$$w^s = \phi(T(G^s),\ L - M^s) > \phi(T(G^b),\ L - M^b) = w^b. \qquad (10)$$

The possibly higher tax and the smaller membership of the trade union both tend to make $w^s > w^b$.

The comparison of the two equilibria is graphically depicted in Fig. 3. The width of the box corresponds to the economy's total work force L, which in a stationary equilibrium is divided into union members (measured from the left to the right) and government employees (measured from the right to the left). The slope of the function is $\phi_M - T_G\phi_T < 0$; it hence takes both the above mentioned effects into account. It is clear that the schedule is steeper the larger the sensitivity of the tax to an increase in the size of the public sector.

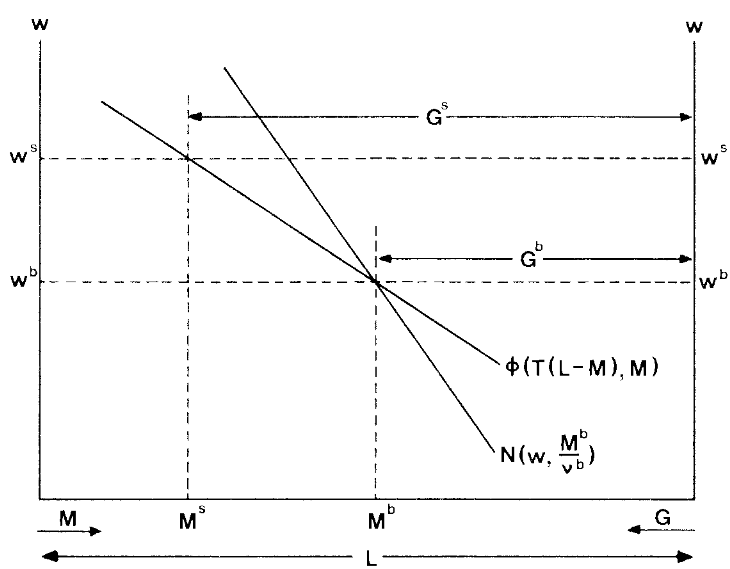

Figure 3

We now turn to a comparison of unemployment in the two economies. It may seem self-evident, at a first glance, that average unemployment is lower in the economy with a larger public sector. There will be fewer private sector workers to employ. Given the wage, there will be less unemployment in every productivity state, and average unemployment should hence be lower in this economy. However, changes in the two determinants of the voted wage studied above, make the comparison of the two economies less self-evident.

First, a smaller size of the private sector implies a smaller union. A worker with higher most-preferred wage will become the new median voter. Second, a larger public sector may lead to a higher tax on workers, causing these to seek compensation by increasing the wage. With a higher wage in the economy with a larger public sector, there is less demand for labor in any productivity state; this tends to create *more* unemployment in economy s. It is hence not clear whether or not the economy with the larger public sector has more unemployment; to get any further we have to resort to some algebra.

Let Y denote the number unemployed workers. The actual number of unemployed depends on the realized productivity level - there will be unemployment when the productivity level falls short of M/ν. That is,

$$Y = M - N(w\ ,\ \alpha) = M - \alpha\nu(w) > 0 \quad \text{for} \quad \alpha < \frac{M}{\nu(w)}\ . \qquad (11)$$

The expected number of unemployed workers is therefore

$$E(Y) = \int_{1-\delta}^{M/\nu(w)} Pr(\alpha)(M - N(w\ ,\ \alpha))d\alpha =$$

$$= \frac{1}{2\delta}\int_{1-\delta}^{M/\nu(w)} (M - \nu(w)\ \alpha)\ d\alpha > 0, \qquad (12)$$

which is unambiguously positive since the minimum value of $M - \nu(w)\alpha$ is zero.

If we differentiate equation (12) with respect to the size of the public sector, we get

$$\frac{dE(Y)}{dG} = -\frac{1}{2\delta}\int_{1-\delta}^{M/\nu} [1 - \alpha\nu_w(\phi_M - T_G\phi_T)]d\alpha \ , \qquad (13)$$

which generally is of ambiguous sign, as was also argued intuitively above.

Equation (13) is easier to interpret when rewritten as;

$$\frac{dE(Y)}{dG} = -\frac{1}{2\delta}\int_{1-\delta}^{M/\nu} \left(1 + N_w(w, \alpha)\ \frac{dw}{dG}\Big|_{\phi}\right) d\alpha \qquad (14)$$

The sign of the integrand is hence determined by the relation between the slope of the labor demand function $1/N_w$, and the slope of reaction function $\frac{dw}{dG}|_{\phi}$. The integrand takes on its minimum value at the upper integration limit M/ν, where the sign is determined by the direction of the inequality:

$$\frac{1}{\frac{M}{\nu}\ \nu_w} \gtrless \frac{dw}{dG}\Big|_{\phi}.$$

The L.H.S of this inequality is the slope of the labor demand function for the productivity level that results in full employment, for given M and w. This function, which we denote the "full employment" function (FE), is formally given by $N(w, \alpha) = M$, evaluated at α equal to a constant M/ν. If the FE-function is at least as steep as the ϕ-function, as illustrated in Fig. 3, then the integrand is positive, and hence $dE(Y)/dG$ is definitely positive.

A sufficient condition for this is that the union is of the stable type, since in this case the PP-function is steeper than the reaction function, while the FE-function is always steeper than the PP-function. Another sufficient condition is that taxes are the same in both economies, i.e. that $T_G = 0$. In this case the minimum value of the integrand is:

$$1 - \frac{M}{\nu}\,\nu_w\phi_M = \left[1 - \frac{A}{2A + \left(1-\tau\right)^2\left(1 + \delta - \frac{M/2}{\nu}\right)V_{cc}}\right] < 0, \qquad (15)$$

where

$$A \equiv (1 - \tau)(1 + \delta)\,\frac{\nu_w}{\nu}\,V_c < 0.$$

Hence, *the expected unemployment rate is unambiguously lower in the economy with the larger public sector, if the union is of the stable type, or if the tax on workers is the same in both economies.* The expected unemployment rate can obviously also be lower even though the tax is higher in economy s, but since the magnitude of T_G is unknown we cannot say anything general about this case. However, it is clear that *if the tax on workers is sufficiently sensitive to the size of the public sector, the economy with the larger public sector might have a higher expected unemployment rate in the case of an unstable union.*

The continuous part of the Y^sY^s-curve in Fig. 4 gives resulting unemployment levels for various α:s, in the economy with the larger public sector. The position of the Y-function is affected in two ways by a decrease in G. First, it is unambiguously steepened. Second, its intersection with the α-axis moves in an ambiguous direction. However, note that

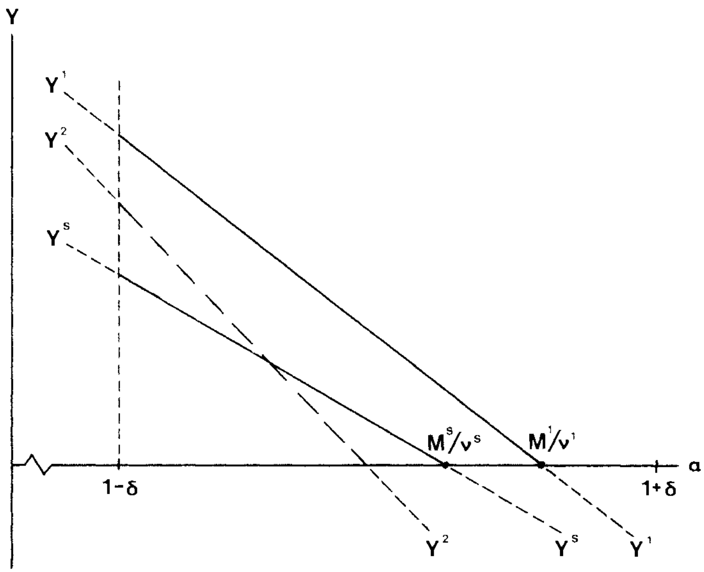

Figure 4

$$\frac{d(M/\nu)}{dG} = -\frac{1}{\nu}\left(1 - \frac{M}{\nu}\,\nu_w(\phi_M - T_G\phi_T)\right) , \qquad (16)$$

and that a sufficient condition mentioned above for $dE(Y)/dG \leq 0$ was exactly that the expression (16) was non-positive. This sufficient condition therefore amounts to the requirement that the proprtion of states with unemployment (i.e. $[(M/\nu) - 1 - \delta]/2\delta$) should not increase with the size of the public sector. It is clear from Fig. 4 that in this case, with the intersection between the new YY-curve and the α-axis to the right of M^s/ν^s, and with the new curve being steeper, the economy with the smaller public sector will have more expected unemployment (the Y^1Y^1-curve). This was also ensured if the tax was constant, since in this case $d(M/\nu)/dG > 0$, as is clear from (13) and (16). However, with a variable tax the economy with the smaller public sector might be represented by Y^2Y^2, in which case the relative expected unemployment level is ambiguous.

5. DYNAMICS OF WAGES AND UNEMPLOYMENT WHEN THE PUBLIC SECTOR CONTRACTS

In the previous section, we studied how the stationary equilibrium depended on the size of the public sector, by comparing the stationary equilibria of two economies b and s, for which $G^b < G^s$. In this section we consider dynamic responses to a reduction in the size of the public sector. But let us before we proceed, specify what is meant by two expressions subsequently used. After the wage is set, the productivity level is realized. Then either the industry's demand for labor is less than the size of the union's membership, and some union members will be unemployed, i.e. there is "excess supply" of union workers. Or, there is "excess demand" for workers, which occurs when the size of the union's membership is not large enough to meet the industry's demand for labor[16].

Assume now the public sector is once-and-for-all reduced from G^s to $\tilde{G}$. Suppose also, for convenience, that the tax rate is kept constant despite the reduced government expenditures. $G^s - \tilde{G}$ workers are hence looking for a private sector job, and some or all of these will be employed in states with excess demand for union workers[17]. If more workers are laid off by the government than the largest possible excess demand, the expected change in the size of the membership ΔM is given by:

$$E(\Delta M) = \int_{M/\nu(w)}^{1+\delta} Pr(\alpha)(N(w, \alpha) - M)d\alpha = \frac{1}{2\delta}\int_{M/\nu(w)}^{1+\delta} (\alpha\nu(w) - M)d\alpha > 0. \tag{18}$$

Since $\alpha\nu(w) - M$ is positive in the whole interval (except at $M/\nu(w)$, where it is zero), $E(\Delta M) > 0$. It is clear that if the number of laid off workers is less than the largest possible excess demand, we also expect workers to enter the union. The main difference is that there is now a positive probability for that all ex-government employees immediately become union members. Hence *the union's membership is expected to increase*. The unambiguity of this result of course stems form the assumption that workers are assumed to remain in the union while temporarily laid off, while they enter the union in high productivity states[18].

Movements of union membership and wages outside a stationary equilibrium are heuristically illustrated by arrows in Figures 1 and 2. The vertical arrows indicate that for a given size of the union membership, the voting procedure will give a wage along the ϕ-function. The horizontal arrows illustrate that the union membership is expected to increase, given the wage. It is clear from Fig. 2 that the unstable union has a tendency to keep increasing as long as there remains unemployed non-union members, given that it initially is larger than M^x. The adjustment to a new stationary state is hence facilitated by instability of the union size. This is not true for an economy with a stable union, which is expected to converge to the size $\hat{M}$ in Fig. 1.

Fig. 5 depicts a possible adjustment path for an economy with a stable union. Assume the public sector is reduced from G^s to $\tilde{G}$ after the wage is chosen in period 1. The realized productivity level in period 1 then happens to be the maximum possible, causing an inflow of $N(w^s, 1 + \delta) - M^s$ new members into the union. But even though the realized productivity was the highest possible, there still remains non-unionized and unemployed workers. The membership, now increased to M_1, votes for the wage w_2 for period 2, which is lower than in the initial equilibrium. In period 2 and 3 the productivity level is low, so there is unemployment also among union members. But, in periods 4 and 5 some workers enter the union. It is possible already in period 4 to have an excess demand for workers large enough to absorbe all the remaining non-union members, but it

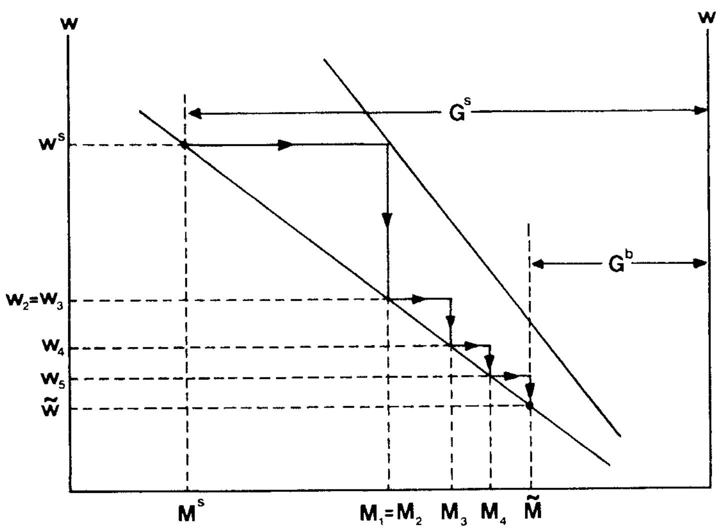

Figure 5

turns out that the realized productivity level does not suffice. In period 5, finally, the productivity level is enough to allow the last workers to become employed, and the economy is eventually in a new stationary equilibrium.

Of crucial interest is, of course, the time required for the economy to reach a new stationary equilibrium. As mentioned above, this is influenced by the relation between the number of workers laid off by the government and the highest possible excess demand for union workers. Another factor is the relation of the absolute magnitudes of G^s and $\tilde{G}$ to M^x or $\hat{M}$, given the PP-function and the ϕ-function. The closer to the intersection the reduction of the size of the public sector takes place, the longer is the adjustment process expected to be. Heuristically, the further into the wedge between PP and ϕ in Figures 1 and 2 the reduction in G occurs, the more periods required, regardless of whether the union is stable or unstable. It is actually possible in the case of a stable union that, before it has grown large enough to encompass all workers not employed by the government, a median voter chooses a wage such that the probability of employment of the least senior worker approaches zero. This situation would arise if the public sector is reduced below the size at which the reaction function and the positive probability function intersects. In this case the probability of excess demand for labor is zero, and the growth of the size of the union has come to a half. The union has acqired its natural, or stable, size. Formally, this occurs if simultaneously $N(w, 1 + \delta) = M$, and w maximizes $U^{M/2}(w)$ for $M < L - G$. This is the most troublesome case, since there will now be created *a pool of permanently unemployed workers;* permanently meaning that no further wage adjustment will reduce unemployment, in this economy.

Summarizing, if the government lays off sufficiently few workers one can not be sure that any unemployment will occur, even if some is to be expected. If it lays off sufficiently many there will definately be temporary unemployment. The size of the union will eventually be the total work force less those employed in public sector, but the time required for the union to absorbe these ex-government employees might, of course, nevertheless be considerable. Finally, if the public sector becomes small enough, it is possible that the unemployment becomes permanent, if the union is of the stable type. It hence seems that in this economy a fear that a reduction in the size of the public sector might have serious consequences in terms of high and persistent unemployment, is not necessarily unwarranted.

6. CONSEQUENCES OF AN EMPLOYMENT STABILIZATION POLICY

The next chapter analyzes an employment stabilization policy that is based on government interventions into the labor market. The basic idea of the policy is that by employing workers in low productivity states, and laying them off in states with higher productivity, the government manages to reduce swings in the level of total employment. Taxes are kept constant and are set at a levels at which the budget is expected to balance in the long run. It is shown that such a policy might have undesired long run allocational effects, even though the policy is constructed in a way which would, under certain assumptions about trade union perceptions, leave the economy's "normal" equilibrium unaffected. The purpose of this section is to show a similar result, but one that stems from entirely different reasons.

For expository purposes we here consider an extreme form of accommodation policy, where the government in periods with unemployment expands the public sector to completely absorb all unemployment, and in higher productivity states lays off enough workers to allow the private sector to expand as it wishes. The government hence acts as an employer of last resort.

Some short and long run consequences of such a stabilization policy are also investigated in Söderström and Viotti (1979). Their focus is on how exogenous changes in tbe (nominal) wage rate affect an economy persuing such an accommodation policy. Here, on the other, the main concern is how such a policy in itself affects the endogenous wage.

That the government absorbs unemployed workers implies, in terms of our model, that they leave the union[19]. The effect of introducing the stabilization policy is therefore to change the assumtion made above regarding asymmetric exit and entry, to a more symmetric one, were workers leave when they are laid off in the private sector and enter when they are rehired.

Our aim is to provide an example, rather than any general results. To make our point as simply as possible, we assume that the trade union reaction function is linear in the size of the membership, and we use the following notation:

$$N(w_t , \alpha_t) = \alpha_t(\beta - \gamma w_t) ; \qquad \beta > 0, \quad \gamma > 0$$
$$\phi(M_t) = \mu - \lambda M_t ; \qquad \mu > 0, \quad \lambda > 0. \qquad (19)$$

Assume that the trade union is of the stable type, and that it in the absence of the stabilization policy has aquired its stable size. The union's size is hence given by the intersection of the PP-function and the ϕ-function. Formally, the stationary state values

$(\hat{w}, \hat{M})$ are given by the solution to:

$$w = \phi(M), \tag{20}$$

and

$$M = (1 + \delta)\ \nu(w) \tag{21}$$

Inserting the parameters, and solving for the wage,

$$\hat{w} = \frac{\mu - \beta\lambda(1+\delta)}{1 - \gamma\lambda(1+\delta)}\ . \tag{22}$$

The denominator is positive, since the union is of the stable type, and since we are only interested in positive wages, we assume that the numerator is also positive[20].

Let us now introduce the employment stabilization policy. The size of the membership that votes for next period wage is, when exit and entry is symmetric, whatever happened to be the private sector employment in the previous period. The dynamic system is hence described by:

$$w_{t+1} = \phi(M_t), \tag{23}$$

and

$$M_t = \alpha_t\ \nu(w_t), \tag{24}$$

and the development of the wage is given by the Markov process:

$$w_{t+1} = \phi(\alpha_t\ \nu(w_t)) = \mu - \lambda(\beta - \gamma w_t)\alpha_t. \tag{25}$$

The conditional expected value of w_{t+1}, given a realized w_t, is:

$$E(w_{t+1}|w_t) = \mu - \beta\lambda + \gamma\lambda w_t. \tag{26}$$

Let $\bar{w}_t \equiv E(w_t)$. The expected value of w_{t+1} is then:

$$\bar{w}_{t+1} = \mu - \beta\lambda + \gamma\lambda\bar{w}_t \ , \tag{27}$$

which is a first-order linear difference equation. That the union is stable implies that $\gamma\lambda < 1$. Furthermore, since we assumed that the initial wage was positive (i.e. that $\mu - \beta\lambda(1+\delta) > 0$), it follows $\mu - \beta\lambda > 0$. The difference equation hence has a positive and stable solution, as is clear from Fig. 6, and the expected wage monotonically converges to:

$$\lim_{t\to\infty} E(w_t) = \bar{w}^* = \frac{\mu - \beta\lambda}{1 - \gamma\lambda} > 0 \tag{28}$$

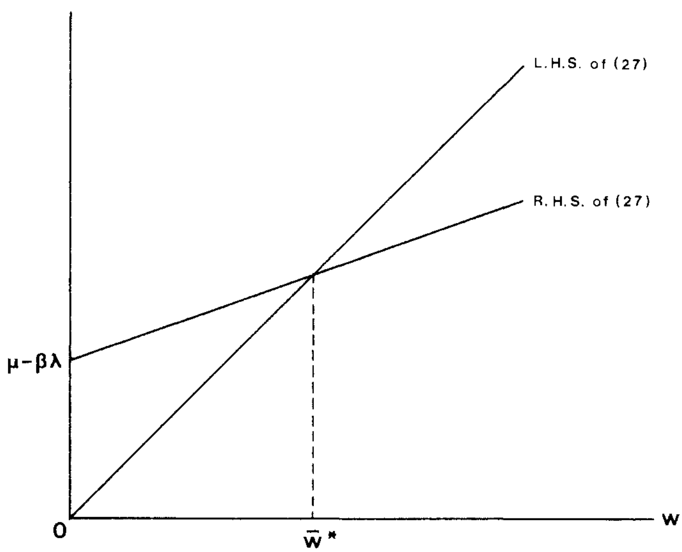

Figure 6

Comparing wages with and without the employment stabilization policy, we find that its introduction increases the expected long run wage[21]:

$$\bar{w}^* = \frac{\mu - \beta\lambda}{1 - \gamma\lambda} > \frac{\mu - \beta\lambda(1+\delta)}{1 - \gamma\lambda(1+\delta)} = w \qquad (29)$$

As a consequence, *the effect of the employment stabilization policy is to decrease the expected long run private sector employment:*

$$E(N(w\ ,\ \alpha)|\text{stab. pol.}) = \nu(\bar{w}^*) < \nu(\hat{w}) = E(N(w\ ,\ \alpha)|\text{no stab. pol.}) \qquad (30)$$

Intuitively, actual employment varies between periods even in the long run, in this economy. If the government strives to counteract these variations by expanding the public sector in times of low productivity, wages will increase due to the reduced size of the union. These wage increases reduce the likelihood of that the productivity levels in the consecutive periods will suffice to absorb workers just hired by the government, if they were to be laid off. There is hence a tendency towards a reallocation of workers from the private to the public sector. The stabilization policy does not only *stabilize* employment, but it also in an average sense *shifts resources* over to the government sector.

7. CONCLUDING COMMENTS

Various factors may in this model contribute to create temporary or permanent unemployment. The most obvious factor is the monopoly power of the union - the union actually manages to set the wage the membership votes for. Another factor was emphasized above in the context of a public sector contraction: the membership of the private sector union does not comprise all those who are looking for a private sector job. Suppose the government accepts the monopoly power of the union in the sense that it accepts the wage the union would choose if the union's membership included all workers not employed by the public sector. Then, what could this government do if it wanted to reduce the size of the public sector, but was anxious to avoid unemployment?

One thing to do would be to *create possibilities and incentives for unorganized workers to join the union,* even though they are presently jobless. Such a policy, leading to a higher unionization rate, would somewhat paradoxically be supported by capitalists, since the lowering of the wage would increase expected profits[22]. But a (marginal) wage reduction would not only be welcomed by capitalists, it would increase the expected utility of union members less senior than the median worker. The policy would obviously be welcomed by the unorganized workers, and probably also by public sector employees, since reduced expenditure on employment benefits and an increased tax base makes possible a reduction of the tax. (The latter reasoning applies also to a group of workers with seniority rankings better than the median.) It thus seems as such a policy in our economy could get a wide support.

But the government might for constitutional or other reasons find it infeasible to interfere with the trade union's way of recruiting members. However, the same end could be achieved by means of a more traditional policy tool - the income tax on workers. An exogenous tax cut creates a situation with, relatively speaking, too much consumption and a too low probability of employment for the trade union. The union will therefore adjust by lowering the wage.

Hence, *by accompaning a contraction of the size of the public sector by an appropriate tax cut, the government can increase the expected number of new members*[23].

An immediate implication of this is that *a contraction of the public sector which is caused by an attempt to close a government budget deficit and which therefore does not lead to a tax cut, will create the "worst" unemployment problems.* Within this model such a policy would be just opposite to what should be done to lessen the unemployment problem.

The policies suggested in this section have associated costs. If we are to evaluate these policies more thoroughly, these costs obviously have to be taken into account. It is obvious that the arguments might have to be modified if the government or the trade union are concerned and aware of e.g. the future financing of a budget deficit.

The analysis showed that one should expect a reduction in the size of the public sector to be associated with more unemployment, than is expected to prevail in the new stationary equilibrium. The reason put forward here is based on differences in interest between workers who differ in seniority. Needless to say, this is only one of several potential sources for this phenomenon; search unemployment, for instance, is another concievable reason.

The potential downward rigidity of the wage might in some circumstances be interpreted as a *ratchet effect* in the relation between the size of the public sector and the real wage. Consider the following example. Assume the size of a stable union corresponds to the intersection of the PP-schedule and the ϕ-schedule. There will then be excess supply of union workers in all productivity states, except for the highest. The public sector can therefore undertake at least some expansion in virtually every state, if the government so wish, and the full wage response to such an expansion would come in the next voting. However, a contraction back to the previous size does not come as easily. To come down to the former wage, all workers laid off by the government have to re-enter the union. But the possibility for this hinges on realized productivity levels. Furthermore, if the "wedge" between the ϕ-function and PP-function is "small", it might be very hard to get everyone into the union. There are increasingly few states with excess demand, the more the unemployment pool is emptied.

This ratchet-type effect has some implications. It is clear that there exists no one-to-one correspondence between the size of the public sector and the wage, instead the wage is partly determined by "history". We might hence be misled if we use econometric estimates of this relationship that are based on periods of public sector expansion, if we are trying to forecast macroeconomic shorter run effects of a public sector reduction[24].

It was assumed above that entrants into the private sector labor market were workers who were laid off by the government. It should be pointed out, however, that other groups of entrants could also face a similar type of problem. For instance, it is often claimed that women are laid off in the downswings of economic activity proportionally more than men. This has a ready interpretation in terms of our model, to the extent that they on average have lower seniority than men. Another such group which comes to mind is school-leavers. But here the case is less clear, since we now should take account of retirements as well.

In sum, this chapter analyzes unemployment in an economy where the demand for labor is determined entirely by the real wage, and where hence Keynesian aggregate demand effects are absent. The main conclusion is that in the economy under study the wage is flexible upwards, but responds sluggishly downward to a cut in the size of the public sector. A reduction in the size of the public sector might, due to this downward rigidity, have severe consequences in terms of high expected unemployment rates, and it is even possible that the economy might end up with a pool of permanently unemployed workers. The downward rigidity of the wage implies also that employment stabilization policies that work through government purchases of unemployed labor, might have long run allocational effects, leading to lower long run private employment. Tax policies, on the other hand, do not suffer from these disadvantages, and have also another advantage in that only tax policies can move the economy away from a situation with permanent unemployment. Important disadvantages with tax policies are that their effects are delayed, and that the effects are contingent on realized productivity levels in ensuing periods.

This chapter was inspired by the observation that even though most smaller developed economies are characterized by strong trade unions and large government sectors, little formal macroeconomic analysis has been pursued about possible linkages between these two phenomena. The purpose was not to provide any generally valid results, but only to suggest possible relationships by means of simple examples. The analysis is obviously extremely partial, and needs to be further developed in a number of ways. First, the analysis is far too atemporal; this would have to be dealt with if e.g. investments were to be included. Second, the type of economies that have inspired the analysis are usually very open; many aspects of wage formation in these countries seem to be best understood in the context of international economics. Third, the wage bargaining process could be more satisfactorily depicted. This might call for the introduction of other unions and an employer's side. Fourth, there is of course a mutual influence between trade unions and governments, which suggests that their relationships might fruitfully be viewed from a game theoretic perspective.

FOOTNOTES

*I am grateful to Gene Grossman and Torsten Persson for many helpful suggestions and discussions. I also wish to thank Lars E.O. Svensson, Jörgen Weibull, Magnus Blomström, John Cuddington, Anders Forslund, Kari Lantto and Assar Lindbeck for valuable help. Comments on earlier drafts were also provided by participants in seminars in the Stockholm Theory Workshop, at the Institute for International Economic Studies, University of Stockholm, at the Graduate Institute of International Studies, Geneva, and at the Stockholm School of Economics. Financial support from the Bank of Sweden Tercentenary Foundation is gratefully acknowledged.

1. The Cambridge economists Jackson, Turner and Wilkinson (1972) argue that

> "...the effective rate of increase of employee living standards depends much more on government decision than on success of union wage demands and the movements of nominal money wages or of prices..." (p. 111).

2. However, see Gylfason and Lindbeck (1982), Driffill (1983), Sampson (1983) and the next chapter for analyses of this kind.

3. For the case of Britain, see e.g. Bacon and Eltis (1978).

4 An example of an organization recommending this type of rule is the Swedish Confederation of Trade Unions (LO).

5. Partial derivatives are denoted by alphabetic subscripts.

6. The reason we do not introduce any explicit government budget constraint is that we would then have to specify how public sector

wages are determined. The tax function (1) captures the essential point in that it allows for the tax to increase in the size of the public sector.

7. Most of the analysis would remain qualitatively unchanged if we assumed $\nu_{ww} \leq 0$. What we need to assume is that ν_{ww} isn't "too" positive.

8. This type of randomness could e.g. stem from a production function $X = F(\alpha K, L)$. In Grossman (1983a, b) uncertainty enters as $X = \alpha F(K, L)$. The advantage with the former representation is that the demand function for labor is somewhat easier to work with, while the main advantage with Grossman's (1983a, b) representation is that the randomness in the labor demand function could also be due to fluctuating goods prices.

9. The reader is referred to the original Grossman papers for a more thorough presentation of the model.

10. See e.g. Atkinson and Stiglitz (1980).

11. Superscripts i, M/2, and M, subsequently denote individuals.

12. The second-order condition is that

$$\Lambda_w = (1 - \tau)^2\left(1 + \delta - \frac{M/2}{\nu}\right)V_{cc} + (1 - \tau)M \frac{\nu_w}{\nu^2} V_c - \\ - \frac{M\ \nu_w^2}{\nu^3}\left[V\left((1 - \tau)w\right) - V(b)\right] < 0,$$

which is always fullfilled.

13. V_c and V_{cc} should be evaluated at $(1 - \tau)w$.

14. A thorough theoretical analysis of the relation between union wages and income taxes is provided by Herzoug (1983).

15. We assume that the public sector wage is such that government employees prefer to remain in the public sector, rather than to line up at the end of the private sector's job queue; that is, that these workers' expected utilities are higher when working in the public sector. This would be the case if public sector workers considered it unlikely that they would be laid off and wages in the two sectors were approximately the same.

16. We use this terminology for convenience. It is obviously slightly misleading - there is, of course, no demand for union members as such, but only for workers.

17. The lucky ones are picked e.g. by random draw.

18. See section 6 for a discussion of an alternative assumption.

19. The seniority ranking is usually not linked to union membership.

20. The definition of a stable union is that the reaction function is less steep than the PP-function:

$$-\lambda > -\frac{1}{(1+\delta)\gamma} \Leftrightarrow \gamma\lambda < \frac{1}{1+\delta} < 1.$$

21. To establish the direction of the inequality, differentiate the expression for $\hat{w}$ w.r.t. $(1+\delta)$:

$$\frac{d}{d(1+\delta)}\left[\frac{\mu - \beta\lambda(1+\delta)}{1 - \gamma\lambda(1+\delta)}\right] = \frac{\lambda\gamma}{(1 - \gamma\lambda(1+\delta))^2}\left(\mu - \frac{\beta}{\gamma}\right).$$

That the union is stable requires that the PP-function's intercept β/γ is larger than the reaction function's intercept μ. The expression is hence negative, and with $(1+\delta) > 1$, $\hat{w} < \bar{w}^*$.

22. The Swedish Employers Federation has explicitly supported a high unionization rate. The alleged reasons have been different from the one suggested here, and instead more concerned with e.g. the likelihood of uncontrolled industrial disputes, etc.

23. This is easily seen formally if we differentiate the expected number of entrants with respect to the tax τ;

$$\frac{dE(\Delta M)}{d\tau} = \frac{dE(\Delta M)}{dw} \frac{dw}{d\tau} = \nu_w \phi_\tau \frac{1}{2\delta} \int_{M/\nu}^{1+\delta} \alpha \, d\alpha < 0.$$

That is, by decreasing the tax the expected number of entrants into the union increases.

24. Some consequences of exogenous ratchet effects are discussed in Calmfors (1982).

REFERENCES

Atkinson, A. and J. Stiglitz, 1980, *Lectures on Public Economics*, McGraw-Hill, Maidenhead, Berks.

Bacon, R. and W. Eltis, 1978, *Britain's Economic Problem: Too Few Producers*, Macmillan, London.

Calmfors, L., 1982, "Employment Policies, Wage Formation and Trade Union Behavior in a Small Open Economy", *Scandinavian Journal of Economics* 84, pp. 245-373.

Driffill, E., 1983, "Can Stabilization Policy Increase the Equilibrium Unemployment Rate?", Discussion Paper in Economics and Econometrics, No. 8312, University of Southampton.

Grossman, G., 1983a, "Union Wages, Temporary Layoffs and Seniority", *American Economic Review* 73, No. 3, (forthcoming).

Grossman, G., 1983b, "International Competition and the Unionized Sector", Seminar Paper 251, Institute for International Economic Studies, University of Stockholm.

Gylfason, T. and A. Lindbeck, 1982, "The Macroeconomic Consequences of Endogenous Governments and Labor Unions", Seminar Paper No. 232, Institute for International Economic Studies, University of Stockholm.

Herzoug, T., 1983, "Union Wage Responses to Tax Changes", mimeo.

Jackson, D., H. A. Turner and F. Wilkinson, 1972, "Do Trade Unions Cause Inflation?", Occasional Paper 36, University of Cambridge.

Lindbeck, A., 1983a, "Interpreting Income Distributions in a Welfare State. The case of Sweden", *European Economic Review*, No. 21 pp. 227-256.

Lindbeck, A., 1983b, "Budget Expansion and Cost Inflation", *American Economic Review*, 73, No. 2, pp. 285-290.

MacDonald, I., and R. Solow, 1981, "Wage Bargaining and Employment", *American Economic Review* 71, No. 5, pp. 869-908.

Oswald, A., 1982, "The Microeconomic Theory of the Trade Union", *Economic Journal* 92, pp. 576-595.

Sampson, A., 1983, "Employment Policy in a Model with A Rational Trade Union", *Economic Journal* 93, pp. 297-311.

Söderström, H., and S. Viotti, 1978, "Money Wage Disturbances and the Endogeneity of the Public Sector in an Open Economy", in A. Lindbeck (ed), *Inflation and Employment in Open Economies, North-Holland.*

CHAPTER VI

EMPLOYMENT POLICIES AND CENTRALIZED WAGE SETTING

CHAPTER VI EMPLOYMENT POLICIES AND CENTRALIZED WAGE SETTING*, **

1. INTRODUCTION

There is an emerging consensus in the literature that the US and European unemployment problems have different origins.[1] It is commonplace to regard US unemployment as basically Keynesian, i.e. as resulting mainly from a deficiency of aggregate demand. The European unemployment problem is instead usually seen more as one of classical unemployment, i.e. a situation where the real wage is set above the market clearing level. The aim of this paper is to develop one possible explanation of this real wage problem that is relevant for the smaller European economies.

Usually the problem of too high real wages is seen in connection with the supply shocks of the 70's. For instance Bruno (1981) and Sachs (1979, 1982) stress that the actual real wage is sticky downwards and that it therefore did not respond to the reduction in the full-employment real wage rate that occurred because of the increase in the real cost of energy and increased competition from the NIC countries. Explanations of the inflexibility of real wages have also been set forth by e.g. Grossman and Hart (1981) and MacDonald and Solow (1981).

However, as pointed out by e.g. Gordon (1982), the assumption of an inflexible real wage does not seem to be empirically true. The problem is rather that the real wage *moves* in a way that is not consistent with market clearing. There has been a long-run tendency for real labor costs to rise relative to output per head

---

* This chapter was written together with Lars Calmfors.

in most European economies (cf. e.g. Sachs (1979) or Nickel (1982)). It thus appears that the European real wage problem is more fundamental than just a failure to adjust to the supply shocks of the 70's.

It is common to relate the tendency for real wages to rise in excess of productivity growth to gradually emerging expectations that government employment policies have reduced the need for moderation in wage settlements (cf. e.g. Sachs (1979) or Söderström and Viotti (1978)). The McCracken report (1977) formulates it in the following way: "During the course of the great post-war expansion those responsible for price setting and wage bargaining in many countries became so convinced that governments could, and would, maintain high levels of demand and employment that they increasingly behaved as if there was no way in which they could price themselves out of markets or out of jobs."

However, most of the theoretical analysis of the links between employment policies and wage (price) formation seems to stress the effects on the cyclical sensitivity of *money* wages and prices rather than on *real wage* developments.[2] Two basic assumptions have usually been that: (1) prices and wages are set in decentralized markets where the individual wage or price setter has no reason to expect his individual decision to trigger off policy responses from the government, and (2) that employment policies take the form of monetary or fiscal policies that increase aggregate demand in the goods market. This focus reflects U.S. conditions well, but it is less satisfactory for several of the smaller European economies such as the Scandinavian countries, the Netherlands and Austria where the real wage problems have been most severe, as shown by Calmfors (1982). In these countries wages are set through centralized bargaining between economy-wide employer federations and trade unions that organize the majority of the labor force. Wage setters therefore have every reason to take into account that their decisions may trigger off policy reactions from the government. Moreover, employment policies in these countries have usually taken the form of direct increases in public sector employment or direct subsidization of private firms. These policies increase total labor demand at each given real wage and do not work by reducing the real wage through price inflation as is the case with policies increasing demand in the goods market. There is thus a direct link between employment policies and the expansion of the public sector in these countries as stressed by Söderström and Viotti (1978) and Calmfors (1982a).

Our analysis focuses specifically on these aspects of stabilization policy in the smaller European economies. We construct a model where the real wage is set unilaterally by *one* all encompassing trade union. The reason for not considering more complex institutional set-ups with several trade unions as in Oswald (1979) or with explicit bargaining between employers and trade unions as in McDonald and Solow (1981) is that we want to focus on the interaction between trade union and government behaviour in a system with centralized wage setting.[3]

The paper consists of *three* parts. The *first* part sets out the basic framework. The subsequent two parts discuss the links between real wage behaviour and employment policies under different assumptions. The *second* part discusses cases where the trade union does not anticipate systematic employment policies and the *third* part shows how the results change when the trade union learns to anticipate such policies.

The assumption of centralized wage setting makes it necessary to incorporate game aspects. The solutions in the second part can be interpreted as Nash equlibria and the solutions in the third part as Stackelberg equilibria. We thus distinguish between situations where the values of government policy parameters are correctly anticipated and situations where also the policy rule governing the values of these parameters are foreseen.

The idea of analyzing stabilization policy in game theoretic terms lies close to the rational expectations framework as stressed e.g. by Buiter (1980), Goldfeld (1982), and Kydland and Prescott (1977). In decentralized markets with many small private agents it is natural to regard these as non-dominant players (followers) and the government as the dominant player (leader). But in a centralized system this assignment is less evident (cf. e.g. Faxén (1957) or Johansen (1977)). Indeed we analyze the opposite assignment with the trade union acting as the leader and the government as the follower with its hands tied by an automatic employment policy rule.

2. THE MODEL

Consider an economy with two sectors: a private sector producing a private good (manufactures) and a government sector producing a government good (services). Both sectors use labor, which is variable, homogeneous and mobile between the sectors.

All other factors are fixed. There are two types of agents: capitalists and workers. Capitalists receive all profits in the private sector as their only income. Workers receive only labor income when employed and only a government-paid unemployment benefit when unemployed.

Since the analysis focuses on real variables, money is not introduced. The model may either be interpreted as a model of a small open economy or of a closed economy. In the former interpretation the government good corresponds to a non-traded good and the private good to a traded good, the price of which is exogeneously given from the world market.

2.1 *The Private Sector*

The private sector consists of perfectly competitive firms that maximize profits given the price of private goods (which is set at unity) and the wage rate w. In addition firms also have to take into account the productivity level α which varies around a normal value α_N. This gives a simple representation of exogenous shocks that may motivate stabilization policy. Provided that firms maximize profits individually and thus do not collude, we have a labor demand function $N = N(w, \alpha)$, where $N_w < 0$ and $N_\alpha > 0$. For convenience we also assume $N_{ww} = N_{w\alpha} = 0$.

2.2 *The Government Sector*

The government sector produces services with a unitary input requirement of labor per unit of output.[4] The number of government sector employees G hence refers both to the input and the output of the sector. Public services are distributed to all individuals in equal fixed quantities and at a zero price.

Workers in the government sector are paid the same wage as in the private sector. The government also pays out a fixed untaxed unemployment benefit b to those unemployed (the total number of workers M less those employed E), so total public sector expenditures S are[5]

$$S = G + [M - E(w, G, \alpha)]\, b = S(w, G, \alpha) \qquad (1)$$

$$\text{where } E(w,G,\alpha) = N(w,\alpha) + G \qquad (2)$$

is total demand for labor, i.e. the total number of employed workers.

Taxes are paid both by employed workers and capitalists. Capitalists together pay T^c. Each employed worker pays τ, which is composed of a fixed component τ_o and a variable component that depends upon government expenditure per employed:[6]

$$\tau = \tau_o + \mu \frac{S}{E} = T(\mu, w, G, \alpha) \qquad 0 \leq \mu \leq 1 \qquad (3)$$

The reason for choosing this specific form is that we want to have a simple representation where employed workers pay for a certain share μ of any increase in government expenditures.

We shall not in general specify whether there is budget balance or not. As long as we assume that no intertemporal considerations are made, it makes no difference whether a short-fall between government expenditure and taxes on workers is financed by a tax on capitalists or results in a government budget deficit. A government budget deficit or surplus must have a real counterpart in the goods market. Since all incomes are assumed to be consumed, total private consumption amounts to $Q + wG + (M - E)\ b - T^c - \tau E$, which will differ from output of the private good Q as soon as the government budget is not balanced.

There are two possible interpretations of this. If we regard our model as a model of a small open economy, the counterpart of a budget deficit will be a trade balance deficit, with net imports making up for any excess demand for domestic tradables. The government - as opposed to the private sector - is then assumed to have the opportunity of issuing debt abroad as in Helpman (1977) or Calmfors (1978).

The second interpretation is to regard our model as describing a closed economy. One then has to assume that the government runs a buffer stock of private goods, which is decreased or increased depending upon whether the government budget shows a deficit or a surplus.

2.3 *The Individual Worker*

An individual worker derives utility from consumption of both private and government goods. We assume that the utility function is identical across individuals, and additively separable in the two goods. Labor in itself is not a source of utility or disutility. We take the number of hours each individual works as institutionally fixed, so that when employment (in hours) is below the market-clearing level, a worker is either employed or unemployed. Given the level of (un)employment the ones to be unemployed are singled out by a random draw. The probabilities of being employed and unemployed therefore equal the shares of employed and unemployed in the total labor force. A worker spends all his income on private consumption c. The private consumption of an employed worker therefore equals after-tax income $w - \tau$ and that of an unemployed worker the unemployment benefit b. The expected utility of an individual worker is then:[7]

$$U = \frac{E}{M}\,[V(w - \tau) + Z(G)] + \frac{M - E}{M}[V(b) + Z(G)] =$$

$$= U(w - \tau, G, E), \tag{4}$$

with $V_c > 0$, $V_{cc} < 0$, $Z_G > 0$, and $Z_{GG} < 0$.[8]

2.4 *The Trade Union*

All workers are members of the same trade union. Like Negishi (1979) and Oswald (1982a) we assume that the union acts as a monopolist in the labor market, and sets the wage so as to maximize the unweighted sum of expected utilities of the workers, i.e. to maximize equation (4).[9]

The trade union's preferences can (given τ and G) be represented by a set of indifference curves in the wage-employment plane as in Fig. 1. We make the standard assumption that $w - \tau > b$ so that the utility of an employed worker is always larger than the utility of an unemployed worker; otherwise all workers would prefer to be unemployed. Then the indifference curves are negatively sloped and convex towards the origin, as shown by Oswald (1982a) and Calmfors (1982b).

Given G, we can also draw a labor-demand schedule EE in Fig. 1. The schedule is kinked at B, where the wage becomes so high that all demand for labor from the private sector disappears, and only the demand from the government sector remains. Employment will be demand-determined as long as the demand for labor is lower than the full-employment level M.

If the trade union takes τ and G as independent of its own actions it solves its maximization problem by equating the marginal rate of substitution between wage and employment with the slope of the labor demand curve. If indeed government demand for labor would remain unaffected by *any* wage increases, the trade union would obviously choose an infinitely high wage rate driving the private sector out of existence. We rule out this case and restrict the analysis to cases where both sectors are producing.[10] In Fig. 1 the optimum is thus given by the tangency point A.[11]

In general one cannot rule out the possibility that the trade union's optimization results in a wage that coincides with the full-employment wage (cf. Calmfors (1982b) or Oswald (1982a)). This may be either an interior or a corner solution. Since our aim is to analyze unemployment, we shall follow the literature and restrict the analysis to initial situations like A. As pointed out by Oswald (1982a), this is also the situation in which a trade union is likely to exist; if the trade union would set the market-clearing wage, there is in terms of the model no utility gain for workers from forming a trade union.

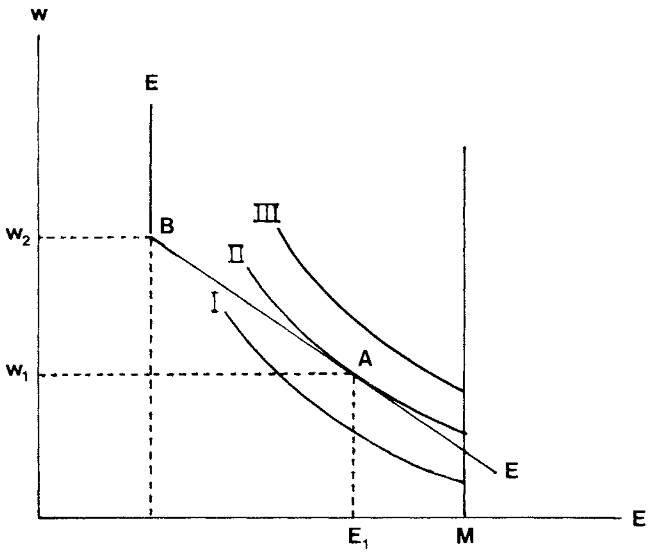

Figure 1

2.5 *Government Employment Policies*

In the following a specific time structure is imposed. It is assumed that the trade union has to set the wage for the next period in advance on the basis of its expectations. The government always takes its decisions after the trade union. We assume that this advantage makes it possible for the government to act on the basis of perfect information.

One way of modelling government behavior is to postulate a utility function for the government and then to derive policy rules from explicit optimization (cf. e.g. Gylfason and Lindbeck (1983)). But our aim is *not* to explain government behavior but to analyze the consequences for trade union behaviour of certain government policies. We shall therefore instead directly postulate a government reaction function, which we find descriptive of post-war employment policies in most smaller European economies. We interpret these policies as unconditional in the sense that attempts have been made to achieve given employment goals more or less independently of the origins of unemployment problems.

The government is assumed to determine the level of government output and employment in the following two-step fashion. In a *first* step it determines a normal level for the size of the government sector, $\tilde{G}$. But government employment is held at this level only if the employment target $\tilde{E}$ of the government is achieved. If employment threatens to deviate from this target, government employment is in a *second step* expanded above or reduced below the normal level G. Following Calmfors (1982b) we assume a simple "leaning-against-the-wind" policy rule, according to which a certain fraction of the deviation from the employment target is eliminated, i.e:

$$G = \tilde{G} + \gamma(\tilde{E} - \underline{E}) \qquad 0 \leq \gamma \leq 1 \tag{5}$$

where $\underline{E} = E(w, \tilde{G},\alpha) = N(w,\alpha) + \tilde{G}$ (6)

is the employment level that would result without employment policies, when government employment is fixed at $\tilde{G}$.

The stipulated policy rule gives a simple representation of employment policies, where the parameter γ measures the degree of "activism".

Using equations (2), (5) and (6), we get:

$$E = \gamma\tilde{E} + (1-\gamma)[N(w,\alpha) + \tilde{G}] = \gamma\tilde{E} + (1-\gamma)\underline{E}(w,\alpha,\tilde{G}) \tag{7}$$

The employment stabilization rule thus makes actual employment E as weighted average of the government employment goal $\tilde{E}$ and the employment $\underline{E}$ that would arise without government interventions.

The government policy rule is illustrated diagrammatically in Fig. 2. The *EE*-line represents the employment function with no government interventions. The policy rule pivots the employment schedule around the point at which $\underline{E} = \tilde{E}$. If $\gamma = 1$ all variations in employment are eliminated and actual employment is pegged at the target level $\tilde{E}$ (the vertical line).

2.6 *Plan of the Analysis*

The analysis below is devoted to the question of how the employment policy rule affects the normal equilibrium of the economy, i.e. the equilibrium when the productivity level is

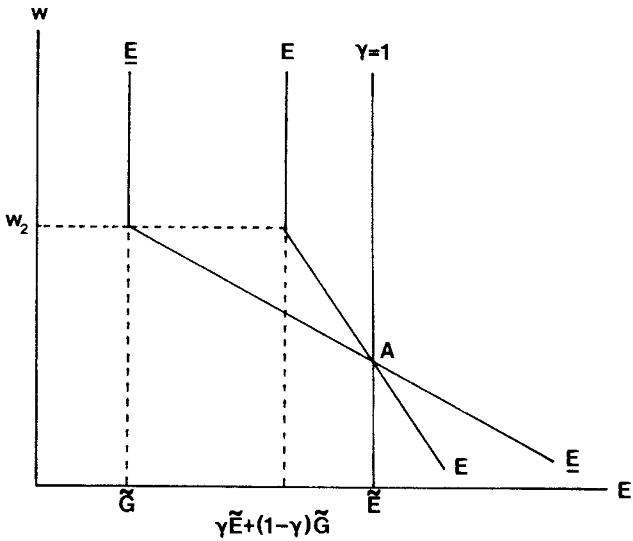

Figure 2

at its normal value ($\alpha = \alpha_N$). We shall vary the assumptions about (1) trade union perceptions of the policy rule; (2) the employment goal of the government and (3) the response of the tax on workers to changes in government expenditure. Formally we shall consider an initial situation without active employment policies, i.e. when $\gamma = 0$ and thus $G = \tilde{G}$, and then study how an introduction of a policy rule changes the initial equilibrium.

3. NASH CASES

This section analyzes situations where the trade union does not perceive the government policy rule and instead regards the size of the public sector as exogenously given. In game-theoretic terminology this situation corresponds to a Nash equilibrium. Formally, the trade union's general optimization problem is to[12]:

$$\max_{w} \quad U(w - \tau, G, E) \tag{8}$$

$$\text{s.t.} \quad G = \text{const},$$

$$E = N(w) + G,$$

and

$$\tau = T(\mu, w, G).$$

The first order condition for an interior utility maximium then is[13]:

$$\Omega(w, G, \mu) \equiv (1 - T_w)U_c + N_w U_E = 0. \tag{9}$$

Assume that the trade union considers a unit's wage increase. $(1 - T_w)$ measures the *after-tax* wage change, which is smaller than the before-tax wage increase (if $\mu > 0$ the per capita tax has to increase to compensate for the increased wage bill in the government sector, as well as for the loss of tax revenues and increased costs for unemployed benefits due to the fall in employment). We have to assume that $T_w < 1$ for an interior solution to exist. The first term in (9) then gives the utility gain from increased private consumption for those who remain employed. The second term captures the utility loss from the decrease in employment. In an optimum these two counteracting effects must balance.

In order to analyze the effects of the employment policy rule we shall need to know how an exogenous change in the size of the public sector affects the wage rate and employment. To see this, equation (9) is differentiated totally holding μ constant:

$$\frac{dw^*}{dG} = - \frac{\Omega_G}{\Omega_w} \tag{10}$$

where $\Omega_G = - (1 - T_w)T_G U_{cc} + (1 - \mu)U_{cE}$

Since we have assumed $T_w < 1$, it follows that $\Omega_G > 0$. It can easily be shown that $\Omega_w < 0$, which is also the second order condition for a maximum (cf. the appendix). An exogenous increase in the size of the public sector thus increases the wage.

The effect on total employment is obtained by differentiation of equation (2) and substituting in the expression for $\frac{dw^*}{dG}$ from equation (10):

$$\frac{dE^*}{dG} = \frac{\partial E}{\partial G} + \frac{\partial E}{\partial w} \cdot \frac{dw^*}{dG} = 1 - N_w \frac{\Omega_G}{\Omega_w} = \frac{1}{\Omega_w} (\Omega_w - N_w \Omega_G) \tag{11}$$

where $\Omega_w - N_w \Omega_G = (1 - T_w)(1 - T_w + N_w T_G)U_{cc} + (1 + \mu)N_w U_{cE} \gtreqless 0$

An exogenous increase in the size of the government sector generally has an ambiguous effect on total employment. This is because the trade union might seek to compensate an increase in the tax due to a larger public sector by increasing the wage. This wage increase lowers private employment, and if it is large enough, total employment might fall.

To analyze the effect of the employment policy rule, we calculate $\frac{dw^*}{d\gamma}$ and $\frac{dE^*}{d\gamma}$. If the signs of these expressions are unambiguous for values $0 \leq \gamma \leq 1$, we also know unambiguously how the equilibrium of the economy under a regime of systematic employment policies deviates from the equilibrium under a regime without employment policies, i.e. when $\gamma = 0$ and thus $G = \tilde{G}$.

In the Nash case it holds that

$$\frac{dw^*}{d\gamma} = \frac{dw^*}{dG}\,\frac{dG}{d\gamma} \tag{12}$$

From equation (5) we get when differentiating totally:

$$\frac{dG}{d\gamma} = (\tilde{E} - \underline{E}) - \gamma N_w \frac{dw^*}{d\gamma}. \tag{13}$$

Together with equation (12) this gives:

$$\frac{dG}{d\gamma} = \left(1 - \gamma N_w \frac{\Omega_G}{\Omega_w}\right)^{-1} (\tilde{E} - \underline{E}). \tag{14}$$

The effect on the wage and employment is now obtained using (10), (11), and (12):

$$\frac{dw^*}{d\gamma} = - \frac{\Omega_G}{\Omega_w - \gamma N_w \Omega_G} (\tilde{E} - \underline{E}), \tag{15}$$

where $\Omega_w - \gamma N_w \Omega_G = (1 - T_w)(1 - T_w + \gamma N_w T_G) U_{cc} +$

$$+ (2 - \gamma(1 - \mu)) N_w U_{cE} \gtrless 0,$$

and $$\frac{dE^*}{d\gamma} = (\Omega_w - \gamma N_w \Omega_G)^{-1} (\Omega_w - N_w \Omega_G)(\tilde{E} - \underline{E}). \tag{16}$$

In the subsections below we will show how the effects of the policy rule under Nash assumptions depend upon how the government-employment target is set and on whether wage earners have to pay for increased government expenditure or not.

3.1 *Government and Trade Union Employment Goals Coinciding*

We start with the case where the initial employment level that would occur without a policy rule ($\gamma = 0$) coincides with the government employment target, i.e. $\underline{E} = \tilde{E}$. This is a situation when the government accepts that the trade union uses its monopoly position to set a wage rate that results in unemployment.

$\underline{E} = \tilde{E}$ gives directly from equations (15) and (16) that $\frac{dw^*}{d\gamma} = \frac{dE^*}{d\gamma} = 0$. The policy rule will thus in this case have no effect on the normal equilibrium of the economy.

This case is most easily illustrated diagrammatically if we assume a constant tax, i.e. $\mu = 0$. In Fig 3. A represents the

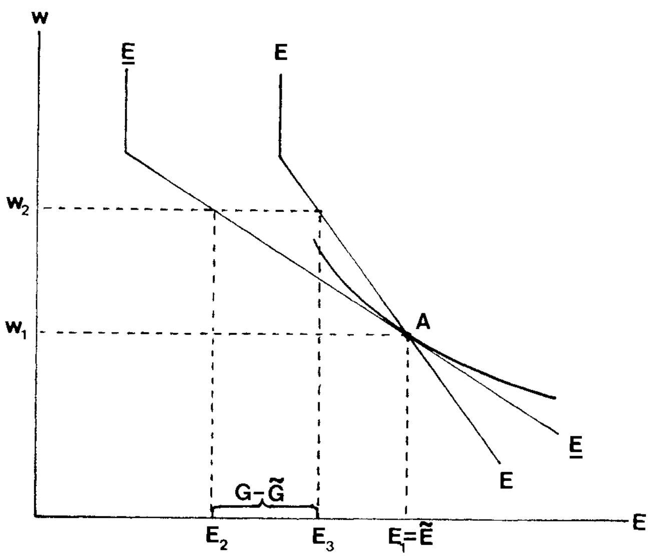

Figure 3

initial equilibrium without a policy rule. The introduction of the policy rule pivots the actual employment curve clockwise around the initial equilibrium point. But if the trade union does not perceive this, it has no incentive to change its actions. Nor has the government, since its employment target is fulfilled if the trade union does not change its wage decision. The point A therefore represents a Nash equilibrium in which the trade union correctly anticipates the level of government employment, even though it does not know the government policy rule.

The Nash equilibrium can also be illustrated with the help of reaction functions as in Fig.4. Equation (10) gives the slope of the trade union reaction function, which is positive (cf. above). The government's reaction function is also positively sloped, since a higher wage triggers off an expansion of the government sector. Differentiation of equation (5) shows that the slope is $\left(\frac{dw}{dG}\right)_G = -\frac{1}{\gamma N_w}$. If the tax is constant, it can easily be shown that the trade union reaction function is steeper (cf. the appendix). If the wage rate chosen by the trade union in the absence of a policy rule (when $G = \tilde{G}$) results in an employment level that coincides with the employment target of the government, the two reaction functions intersect at A along the horizontal $\tilde{G}$-line. Consequently the initial equilibrium will not be disturbed by the introduction of the policy rule.

The employment stabilization rule enters the picture only when the economy is exposed to a shock. Assume for instance that

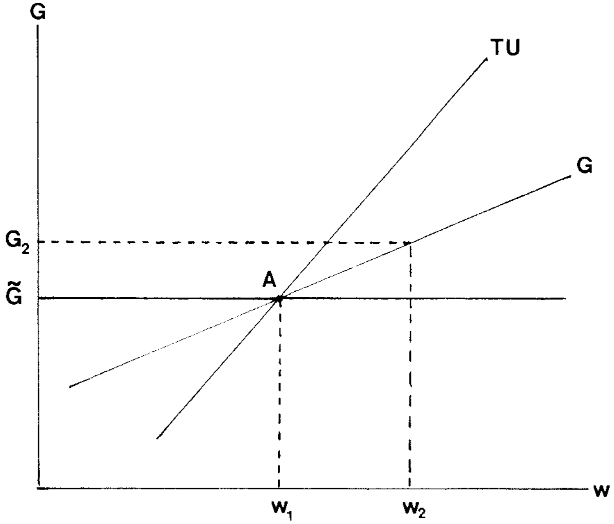

Figure 4

the trade union anticipates a productivity increase (which would shift the labor demand schedule to the right) that does not come about. It therefore sets the wage w_2 in Fig 3, a wage that *ex post* proves too high. Without the employment stabilization rule employment would fall to E_2. But with the policy rule, government employment is expanded and employment falls only to

E_3. In the same way the policy rule reduces employment variations also when actual (unanticipated or anticipated) productivity shocks take place.

Under the assumptions in this subsection the employment policy rule thus works as a pure stabilization policy that reduces employment variations in the case of shocks but that does not affect the normal equilibrium of the economy.

3.2 *Different Employment Targets, a Fixed Tax*

Next we consider the case where the government aims for a higher employment level than the initial one, i.e. $\tilde{E} > \underline{E}$. We also assume that the per capita tax on workers remains unaffected, i.e. that $\mu = 0$. Equation (15) and (16), then simplify to:

$$\frac{dw^*}{d\gamma} = - \frac{U_{cE}}{(U_{cc} + (2 - \gamma)N_w U_{cE})} (\tilde{E} - \underline{E}) > 0 \qquad (17)$$

$$\frac{dE^*}{d\gamma} = \frac{U_{cc} + N_w U_{cE}}{U_{cc} + (2 - \gamma)N_w U_{cE}} (\tilde{E} - \underline{E}) > 0 \qquad (18)$$

The employment rule thus has real effects in this case. Both the wage rate and employment increases.

This case is illustrated in Fig 5. An introduction of the policy rule both shifts the actual employment curve to the right and steepens it.[14] But this is not perceived by the trade union. What the trade union perceives is that if it as before sets the wage rate at w_1, the government will employ $(\tilde{E} - E_1)\, d\gamma$ additional workers in the public sector. If the trade union takes this new level of government employment as given, it

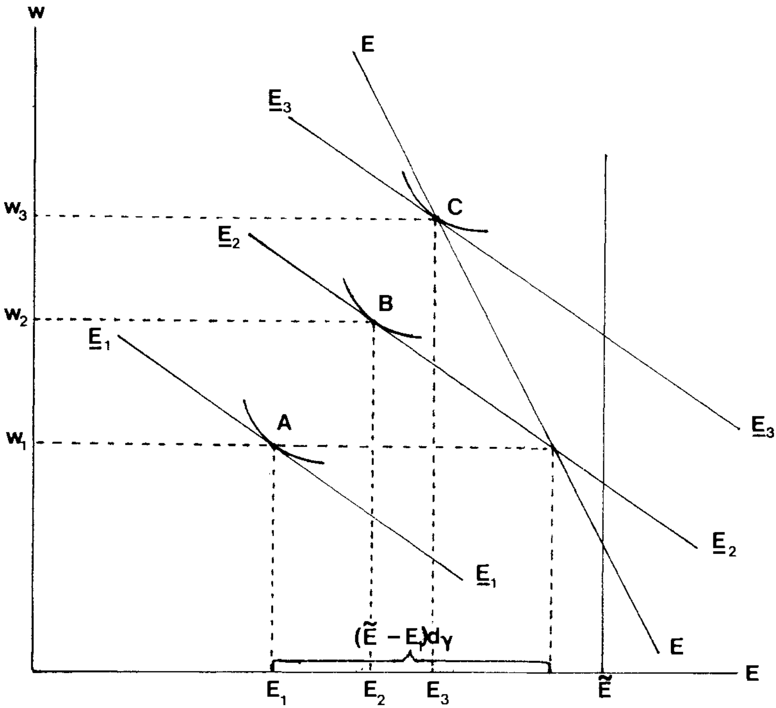

Figure 5

"perceives" only a parallel shift to the right of the initial employment curve. It will therefore choose a new higher wage w_2. This induces a further expansion of government employment. A Nash equilibrium emerges in a point like C at the intersection of the actual and a new perceived employment curve (along which the same level of government employment as in C is taken as exogenous by the trade union).

The new equilibrium can also be illustrated by help of the reaction functions in Fig 6. The trade union reaction function

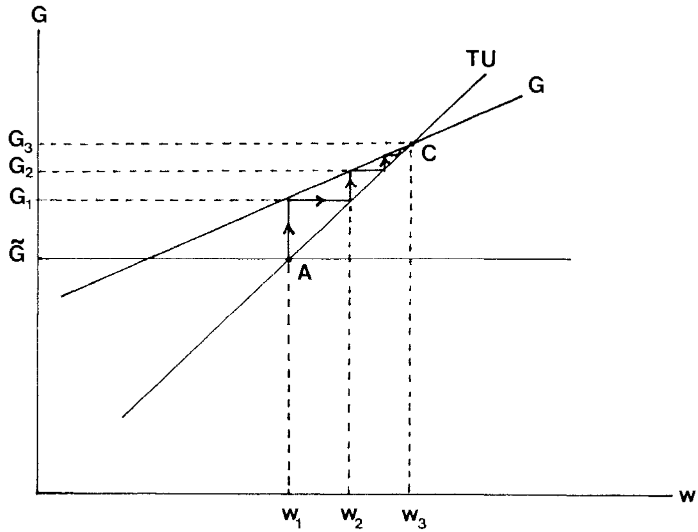

Figure 6

as before passes through A. But if the employment target of the government is not reached at the initial wage w_1, the government will at this wage increase government employment to a level $G_1 > \tilde{G}$. Since the government reaction function is flatter than the trade union reaction function when the tax is fixed, the two reaction functions must intersect in a point like C with both a higher wage and a larger government sector than in A.

We shall not discuss how the transition to this new Nash equilibrium takes place. We merely note that if the trade union has static expectations, i.e. takes the level of government employment prevailing in one period to prevail also in the next, the adjustment process is stable as shown in the diagram.

3.3 *Different Employment Targets, an Endogenous Tax*

Third we consider the case when the government again aims for a higher employment level than the initial one but where also the tax on workers is affected.

In this general case equations (15) and (16) show that the signs of both $\frac{dw^*}{d\gamma}$ and $\frac{dE^*}{d\gamma}$ are ambiguous. The sign of $\frac{dw^*}{d\gamma}$ depends upon the size of γ. For small values of γ the numerator of equation (14) is negative, implying $\frac{dw^*}{d\gamma} > 0$, but for large values it may be the case that $\frac{dw^*}{d\gamma} < 0$. Consequently, we cannot directly infer whether the policy rule drives up the wage rate or not.

Rearranging equation (15) slightly, we find that the sign of $\frac{dw^*}{d\gamma}$ is that of $\frac{1}{\gamma N_w} - \frac{\Omega_G}{\Omega_w}$. But $-\frac{1}{\gamma N_w}$ and $-\frac{\Omega_G}{\Omega_w}$ are the slopes of respectively the government's and the trade union's reaction functions. Hence, as long as the government reaction function is flatter than the trade union reaction function, the new equilibrium will as i Fig. 6 be to the north-east of A. But if γ is so large that the government reaction function becomes steeper than the trade union function, the two curves will intersect to the south-west of A as shown in Fig 7. But this equilibrium will be an unstable one in the sense that if the trade union has static expectations, there will be an explosive adjustment path.

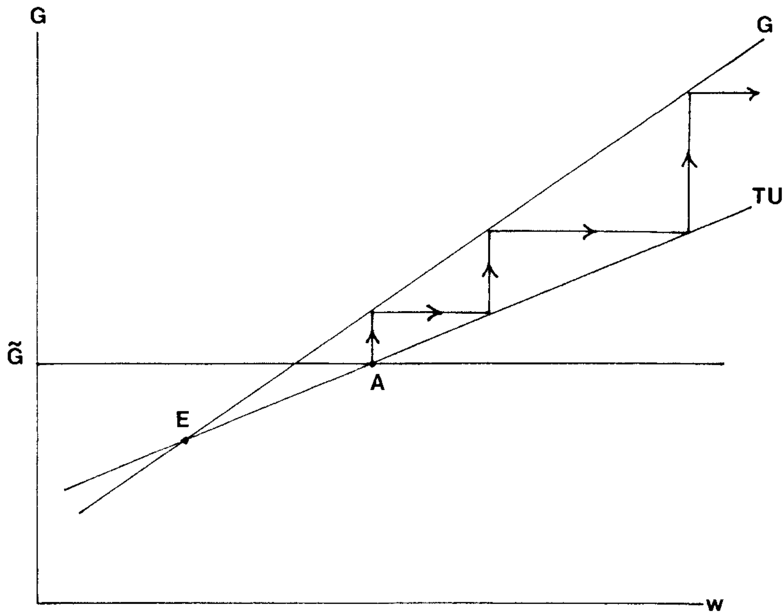

Figure 7

In the stable case it holds that $\frac{dw^*}{d\gamma} > 0$. In this case it is not clear whether the wage rises by more or less than in the case with a constant τ in subsection 2.2. On the one hand the trade union knows that wage increases will increase the tax on workers, which makes it less profitable to drive up the wage (the T_w - term). On the other hand the tax increases that result from the expansion of the government sector reduce the after-tax wage and thus increase the marginal utility of consumption, which tends to drive up the wage (the T_G - term). If the trade union would not perceive the effect of its own wage decision on the tax, the first effect disappears and the wage would clearly rise by more than in the constant tax case. If the tax effects add an additional push to the wage rate, the wage increase may be so large as to cause total employment to fall.

3.4 *Conclusions in the Nash Case*

Although the trade union does not perceive the policy rule determining government behavior, it will have correct expectations about the *values* of government policy variables once a Nash equilibrium is reached. All the same the government policy rule has real effects on the normal equilibrium of the economy as soon as the government employment target deviates from the initial employment level. With a more ambitious employment target the employment policy rule raises both the wage and total employment in the case with a constant tax. With an endogenous tax the wage may even rise so much that total employment is reduced.

4. STACKELBERG CASES

In the Nash case the trade union was not aware of the employment policy rule. In this section we instead consider cases where the union perceives the policy rule. In game theoretic terminology the union behaves as a Stackelberg leader.

Formally the trade union's decision problem in this case is to

$$\max_{w} \quad U(w - \tau, G, E)$$

$$\text{s.t.} \quad \tau = T(\mu, w, G), \tag{19}$$

$$G = \tilde{G} + \gamma(\tilde{E} - \underline{E}),$$

$$\text{and} \quad \underline{E} = N(w) + \tilde{G}.$$

The first-order condition for an optimum now becomes:

$$\Lambda(w, \gamma, \mu) \equiv (1 - T_w)U_c + N_w U_E - \gamma N_w(U_G + U_E - T_G U_C) = 0 \tag{20}$$

The first two terms are the same as in the Nash case. What is new here is the third term. To interpret this term it is helpful to differentiate the trade union utility function (4) with respect to government employment for a given wage:

$$\left.\frac{dU}{dG}\right|_{dw = 0} = U_G + U_E - T_G U_C \tag{21}$$

$U_G + U_E - T_G U_c$ thus measures the effect on trade union utility of one unit's change of government employment. U_G and U_E respectively measure the utility gains from the increased consumption of the government good and the higher employment. $T_G U_C$ measures the utility loss caused by the reduction of the after-tax wage that arises if the increase in government employment causes a tax increase. γN_w in equation (20) gives the change in government employment that is triggered off by a unit's wage change. The third term in equation (20) hence gives the *total* utility effect of the change in the size of the public sector that is caused by a unit's wage change.

The sign of the expression (21) depends on the initial size of the government sector. If the size of the government sector is optimal from the point of view of the trade union given the wage rate and the tax function, $\left.\frac{dU}{dG}\right|_{dw=0} = 0$. If it is less than optimal, $\left.\frac{dU}{dG}\right|_{dw=0} > 0$ and vice versa.

To analyze the effects of the policy rule, we follow the same procedure as before and differentiate equation (20) with respect to w and γ:

$$\frac{dw^*}{d\gamma} = -\frac{\Lambda_\gamma}{\Lambda_w} = \frac{N_w}{\Lambda_w}(U_G + U_E - T_G U_C) - \frac{A}{\Lambda_w}(E - \underline{E}) \gtrless 0 , \qquad (22)$$

where $A \equiv -(1 - T_w + \gamma N_w T_G) T_G U_{CC} - \gamma N_w U_{GG} + (1 - \mu) U_{cE}$.

As can be seen $\frac{dw^*}{d\gamma}$ now consists of two effects. The second one looks very much like the original Nash effect and indeed coincides with this when $\gamma = 0$. The first one will be labelled the Stackelberg effect. The sign of $dw^*/d\gamma$ equals that of Λ_γ, since the second-order condition for a maximum requires that $\Lambda_w < 0$ (cf. appendix).

We also obtain:

$$\frac{dE^*}{d\gamma} = (1 - \gamma)\frac{N_w^2}{\Lambda_w}(U_G + U_E - T_G U_C) + \tag{23}$$

$$+ \left[1 - (1-\gamma)\frac{N_w A}{\Lambda_w}\right](E - \underset{\sim}{E})) \gtreqless 0$$

We shall again investigate how the policy rule affects the economy under different assumptions about the government employment target and tax policies.

4.1 *Government and Trade Union Employment Goals Coinciding, Constant Tax Rate*

First we consider the case where the government employment target coincides with the initial employment level, i.e. $\tilde{E} = \underline{E}$, and where the per capita tax on workers is held constant, i.e. $\mu = 0$. Then only the Stackelberg effects remain in equations (22) and (23) which now reduce to:

$$\frac{dw^*}{d\gamma} = \frac{N_w}{\Lambda_w}(U_G + U_E) > 0, \tag{24}$$

and $$\frac{dE^*}{d\gamma} = (1 - \gamma)\frac{N_w^2}{\Lambda_w}(U_G + U_E) < 0. \tag{25}$$

Therefore, the introduction of the employment policy rule increases the wage and - somewhat surprisingly - *reduces* employment.[15]

The reason for the negative employment effect is that by reducing the opportunity cost of wage increases in terms of lost employment, the policy rule induces the trade union to raise the wage. Since government employment makes up for only a fraction of the employment loss in the private sector, the result must be a fall in total employment. To avoid this, the government would have to even out all variations of employment from its target $\tilde{E}$, i.e. set $\gamma = 1$. But this is an unfeasible solution, since the rational behavior of the trade union would then be to raise the wage rate without limit.

The Stackelberg case is easily illustrated in a reaction function diagram if we plot the trade union's indifference curves (cf. Fig 8). As shown in the appendix, the indifference curves

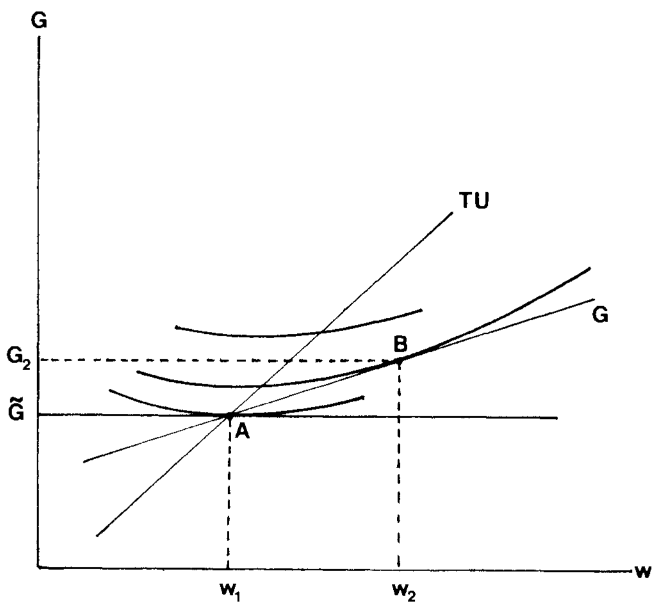

Figure 8

are convex downwards and a higher indifference curve represents a higher level of utility. The trade union will achieve the highest utility level if it chooses a wage rate that makes the government's reaction function a tangent to an indifference curve. In Fig 8 this situation is shown by B with both a higher wage and a larger government sector than in the case without employment policies.

The results can also be demonstrated in the wage-employment diagram (Fig 9). In this diagram the trade union now optimizes

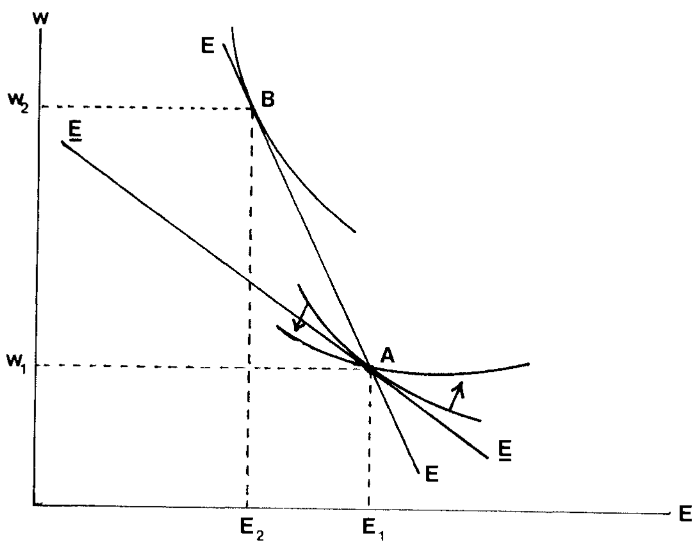

Figure 9

against the steeper employment curve EE, which takes the endogenous responses of the government into account. When drawing the indifference curves we now also have to take into account that the level of government output no longer is regarded as given by the trade union but instead as depending upon the wage rate it chooses. We can therefore draw a new set of indifference curves taking the response of government output into account. They must be flatter than the indifference curves drawn for a given level of government output. The reason is that since a higher wage rate triggers off an increase in government production, a larger decline in employment is now consistent with unchanged utility. Both the steepening of the labor demand curve and the flattening of the indifference curves tend to raise the wage rate. The economy will end up in B.

4.2 *Different Employment Targets, Fixed Tax*

Now consider a case where the tax on workers is still held constant, i.e. $\mu = 0$, but where the government aims at a higher employment than the initial level, i.e. $\tilde{E} > E$. Then we have from equations (22) and (23) that

$$\frac{dw^*}{d\gamma} = \frac{N_w}{\Lambda_w}(U_G + U_E) - \frac{A}{\Lambda_w}(\tilde{E} - \underline{E}) \gtrless 0, \qquad (26)$$

and
$$\frac{dE^*}{d\gamma} = (1-\gamma)\frac{N_w^2}{\Lambda_w}(U_G + U_E) +$$

$$+ \left[1 - (1-\gamma)\frac{N_w A}{\Lambda_w}\right](\tilde{E} - \underline{E})] \gtrless 0 \qquad (27)$$

where $A = U_{cE} - \gamma N_w U_{GG} \gtrless 0,$ (27a)

and $1 - (1 - \gamma)\frac{N_w A}{\Lambda_w} = \frac{1}{\Lambda_w}\Big[U_{cc} + (1 - \gamma) N_w U_{cE} +$

$+ \gamma N_w^2 U_{GG}\Big] > 0.$ (27b)

If $\gamma > 0$ the sign of $\frac{dw^*}{d\gamma}$ is ambiguous. The first term in (26) is positive. But the sign of the second term is ambigous, since the first term U_{cE} in the expression for A is positive, while the second $-\gamma N_w U_{GG}$ is negative. This second term captures the following. If γ is positive, the trade union takes into account that a wage rise increases consumption of government goods. An increase in γ implies more consumption of government goods given the wage. Hence the marginal utility of consumption of government goods is reduced. This creates an incentive to lower the wage rate. The smaller is γ and U_{GG}, the smaller is this effect. Indeed $\gamma = 0$ or $U_{GG} = 0$ are sufficient conditions for this effect to vanish, in which case $\frac{dw^*}{d\gamma} > 0$ unambigously. The effect on employment is, however, still indeterminate, since the two terms in the square bracket in expression (25) still have opposite signs.

The effects in this case are easily illustrated in Fig 10, where we assume that $U_{GG} = 0$ and $\gamma = 0$. The first term in expressions (26) and (27) measures the Stackelberg effect and can be illustrated as the effect of a steepened labor demand schedule (cf. subsection 3.1 above). This effect tends just as

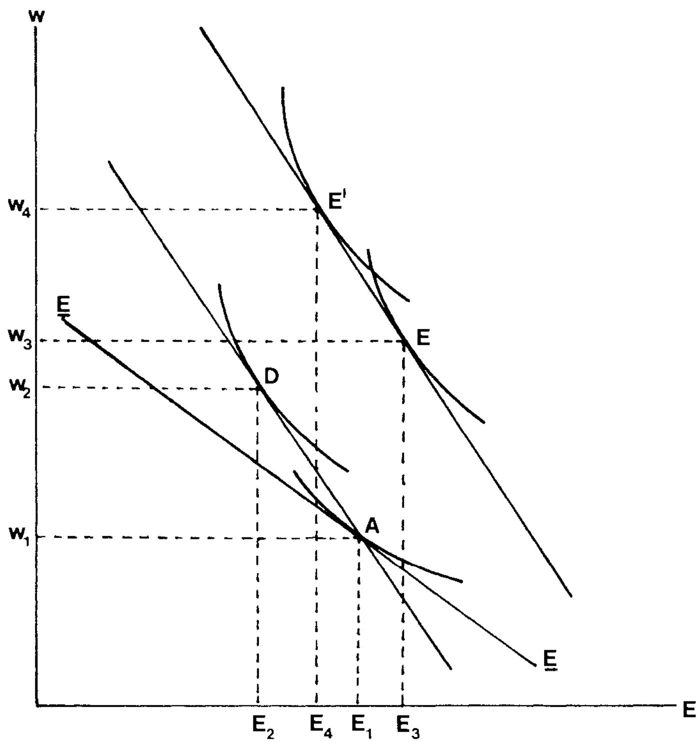

Figure 10

before to increase the wage and decrease total employment (the movement from A to D).

The second term is a Nash effect that can be illustrated as the effect of a parallel rightward shift of the labor demand schedule. This effect tends both to increase the wage and total employment (the movement from D to E or E'). The two effects -- which in this case are analogous to ordinary substitution and income effects - reinforce each other with respect to employment.

Hence the wage must rise, whereas employment may rise or fall.

4.3 *Government and Trade Union Goals Coinciding, Endogenous Tax*

Finally we consider the case when the government employment target again coincides with the initial employment level, i.e. $\tilde{E} = \underline{E}$ but where workers have to bear part of the tax burden of increased government expenditure, i.e. $\mu > 0$. We are then again looking at the pure Stackelberg effect. Equations (22) and (23) then reduce to:

$$\frac{dw^*}{d\gamma} = \frac{N_w}{\Lambda_w} (U_G + U_E - T_G U_c) \gtrless 0, \tag{28}$$

and
$$\frac{dE^*}{d\gamma} = (1 - \gamma) \frac{N_w^2}{\Lambda_w} (U_G + U_E - T_G U_C) \gtrless 0. \tag{29}$$

The signs of both $\frac{dw^*}{d\gamma}$ and $\frac{dE^*}{d\gamma}$ are determined by the sign of $U_G + U_E - T_w U_c = \left.\frac{dU}{dG}\right|_{dw=0}$. Consequently the effect on both the wage and employment depends upon the initial size $\tilde{G}$ of the government sector. If it is smaller than what is optimal for the trade union, $\left.\frac{dU}{dG}\right|_{dw=0} > 0$, and there will be a wage increase and a fall in employment. The reason is that by increasing the wage, the trade union makes a utility gain from the increase in the size of the government sector that is triggered off through the policy rule. If the initial size of the public sector is larger than optimal for the trade union, the opposite result holds.

By comparing equation (28) in the variable tax case with equation (26) in the constant tax case, it is obvious that when the government and the trade union targets coincide, a variable tax will always have a dampening effect on the wage (the $- T_G U_C$ term in equation (28)). The reason is that a variable tax in this case confronts the trade union also with a cost when a wage increase triggers off an expansion of the government sector.

4.4 *Conclusions in the Stackelberg Case*

As soon as the trade union perceives the employment policy rule an additional Stackelberg effect is added to the Nash effect. This Stackelberg effect occurs because the trade union considers how its wage decision affects the size of the government sector. If the initial size of the government sector is regarded as lower than optimal by the trade union, the policy rule gives an incentive for wage increases that in fact reduce employment.[16] Clearly this must happen if the trade union is not faced with any immediate tax increases when the government sector expands and does not consider any future costs. In this case employment policies thus have unambiguously counter-productive effects. If the government aims for a higher level of employment than would result without the policy rule, the effects on employment become ambiguous in the constant-tax case, but the wage rises even more and the private sector contracts even more.

4.5 *Intertemporal Considerations*

In the analysis above we have not specified how changes in government expenditures not covered by taxes on workers are financed. We have left it open whether they cause a government budget deficit or if they are financed through taxes on capitalists. As long as the trade union is myopic this does not matter. But it will, as soon as intertemporal considerations are introduced.

Consider e.g. the case when a government budget deficit today is financed through future tax increases for workers or future reductions of the normal level of government employment. Then the trade union can no longer make the same simple atemporal optimization as above. It must in addition take into account the future welfare losses that are incurred if it sets the wage in such a way that the employment policy rule is triggered off. This case can be analyzed within a two-period framework with an additively separable intertemporal utility function where the constraint is that the government budget has to balance over the two periods. It is straightforward to show that a sufficient condition for the policy rule to raise the wage rate in both periods, if the government and trade union employment goals coincide ($\tilde{E} = \underline{E}$), is that the normal level of government employment $\tilde{G}$ is less than optimal for the trade union in both periods (the mathematics is not reported here).

As long as there are no investments in the model, taxes on capitalists have no future effects. But if a link between capitalists' income and the future capital stock is introduced, the trade union would have to take into account the effects on investment that arise if taxes on capitalists are raised to finance an expansion of the government sector.

The probability that the introduction of a policy rule increases the wage also in an intertemporal framework is, of course, large if the trade union underestimates the future costs of present government expenditures, or believes that these costs can be shifted over to future generations.

5. SUMMARY

One should in general expect systematic employment policies in the form of fiscal policies that shift the labor demand curve to have real effects on the normal equilibrium of the economy. Such policies thus do not only reduce variations in macroeconomic variables but also affect the average values around which these variations take place.

We have shown a menu of cases which differ with respect to the assumptions on (1) how ambitious employment policies are; (2) how changes in government expenditure are financed and (3) trade union perceptions about the policy rule.

Employment policies are purely stabilizing when the government accepts the employment level that results from the trade union's wage decision and when the trade union does not perceive the policy rule. But if the government aims for a higher level of employment and/or the trade union perceives the policy rule, the wage rate is likely to be driven up and the government sector to expand at the expense of the private sector. It is quite possible that the economy ends up in a stagflationary situation where also total employment is lowered.

Tax changes for workers due to changes in government expenditure may have quite different effects under different assumptions about trade union perceptions. If the trade union does not perceive the policy rule, a tax increase on workers to finance an expansion of the government sector tends to increase the wage. But when the policy rule is anticipated, the wage may increase for exactly the opposite reason, namely because the trade union is *not* confronted with increased tax costs when wage increases trigger off an expansion of the government sector.

Our analysis suggests a possible interpretation of the macroeconomic developments in the smaller European economies with centralized wage setting. The Nash cases without anticipations of employment policies could be regarded as applicable to the earlier post-war period, when trade unions had not yet learnt to take employment policies into account. A Stackelberg case where wage earners drive up wages in anticipation of government policy responses would seem to catch important aspects of the combined stagflation and expansion of the government sector in the seventies in these economies. This Stackelberg effect is a consequence of centralized wage setting and could arise in a decentralized system like in the U.S. only if employment policies are selective and thus directed towards the individual industry.

Obviously the analysis here represents only a first step. Several extensions suggest themselves: (1) to examine cooperative as well as non-cooperative solutions, which would make it necessary to introduce a well specified utility function for the government, (2) to introduce several trade unions and discuss the possibility of Prisoner's Dilemma situations and (3) to introduce the employer side as an active agent in the wage setting process and analyze explicit wage bargaining situations.

FOOTNOTES

** We are grateful for comments from participants in the Institute for International Economic Studies Summer Workshop 1982 and from participants in a seminar at the Stockholm School of Economics as well as from Gene Grossman, Thorvaldur Gylfason, Assar Lindbeck, Paul Marginson, Andrew Oswald, Mats Persson, Torsten Persson and Assaf Razin. The research has been financed by grants from Bankforskningsinstitutet, the Nordic Economic Research Council and the Bank of Sweden Tercentenary Foundation.

1. Cf. e.g. Malinvaud (1982) or Sachs (1982).

2. Cf. e.g. Fellner (1980), Phelps (1978) or Sachs (1980).

3. The model here thus is a theory for the real wage that would be desired by wage earners when acting as a collective. It remains to be shown that the actual bargained real wage in a more complex model moves in the some direction.

4. The analysis would not be qualitatively different if the marginal productivity of labor was diminishing.

5. M and b are henceforth suppressed under the function sign.

6. Partial derivatives are given in the appendix.

7. V_c and V_{cc} should always be evaluated at the consumption level $w - \tau$.

8. Partial derivatives are given in the appendix.

9. Cf. also Layard (1982). Alternatively we could, like Grossman (1981) use a median voter approach. But the results would be basically the same, as long as the size of the trade union cannot change endogenously. Note also that our general formulation of the trade union utility function covers the special case when the trade union maximizes the wage bill. It has been shown by Oswald (1982b) that the trade union utility maximization problem then boils down to maximization of the wage bill in the case when the trade union runs its own internal unemployment insurance scheme without any government-paid unemployment benefits.

10. Implicitly we thus impose a restriction on the government's willingness to accommodate wage increases in the government sector.

11. An implicit assumption above is that the choice of which workers will be unemployed is both random *and* repeated frequently. If e.g. the same individuals were to be unemployed period after period, it is hard to see why they would remain in the trade union and refrain from bidding down the wage rate.

12. α is henceforth surpressed under the function signs since we focus on cases where $\alpha = \alpha_N$.

13. As noted above there exists the possibility that such an interior solution does not exist. If it is the case that

$(1 - T_w)U_c + N_w U_E < 0$ for all $0 \leq E \leq M$, the trade union obtains maximum utility by setting the wage at the full-employment level. For the reasons given above we do, however, focus on interior solutions.

14. It also "renumbers" the indifference curves, but because of the separability assumption on the trade union utility fuction it will not change their slopes.

15. Cf. also Calmfors (1982b) where the same result is derived.

16. Since no utility function for the government has been specified, the model does not indicate whether $\tilde{G}$ is likely to be larger or smaller than what is optimal for the trade union. But it does not seem farfetched to conceive of situations where the government takes into account the interests of both capitalists and workers, and where capitalists obtain a proportionalely smaller share of the benefits of an increased government sector than what they pay for through taxes (note that the capitalists get utility only from the output of the government sector but not from public employment). Then one would expect the government to choose a lower $\tilde{G}$ than the trade union desires.

APPENDIX

Partial Derivatives of the Tax Function

$$T_w = \frac{1}{E}[\mu(G - bN_w) - N_w(\tau - a)] > 0$$

$$T_{ww} = -\frac{2}{E} N_w T_w > 0$$

$$T_G = \frac{1}{E}[\mu(w - b) - (\tau - a)] > 0$$

$$T_{GG} = -\frac{2}{E} T_G < 0$$

$$T_{Gw} = \frac{1}{E}(1 - T_w - N_w T_G) - \frac{(1 - \mu)}{E} > 0$$

Partial Derivatives of the Utility Function

Set without loss of generality the size of the union's membership to unity, and let $c \equiv w - \tau$. Then

$$U_c = EV_c > 0; \quad U_{cc} = EV_{cc} < 0; \quad U_G = MZ_G > 0;$$

$$U_{GG} = MZ_{GG} < 0; \quad U_E = V(c) - V(b) > 0 \text{ with } c > b;$$

$$U_{EE} = U_{cG} = U_{GE} = 0; \text{ and } U_{cE} = V_c > 0.$$

Second-Order Condition in the Nash Case

$$\Omega_w = (1 - T_w)^2 U_{cc} + 2N_w U_{cE} < 0 \text{ always.}$$

Slope of Reaction Function in the Nash Case with a Constant Tax

Inserting $\mu = 0$ in eq. (10) we get

$$\frac{dw^*}{dG} = - \frac{1}{\frac{U_{cc}}{U_{cE}} + 2N_w} .$$

The slope of the government's reaction function is

$$\left(\frac{dw}{dG}\right)_G = - \frac{1}{\gamma N_w} .$$

Since $\frac{U_{cc}}{U_{cE}} < 0$ and $0 < \gamma \leq 1$, it follows that

$$0 < \frac{dw^*}{dG} < \left(\frac{dw}{dG}\right)_G ,$$

i.e. the trade union's reaction function has always a steeper slope in (w, G) space than the government's reaction function in this case.

Second-Order Condition in the Stackelberg Case

$$\Lambda_w = (1 - T_w + \gamma N_w T_G)^2 U_{cc} + 2(1 - \gamma)N_w(1 - T_w + \gamma N_w T_G) U_{cE} -$$
$$- (T_{ww} - 2\gamma N_w T_{wG} + \gamma^2 N_w^2 T_{GG}) U_c + \gamma^2 N_w^2 U_{GG}$$

(a) $\gamma = 0$ implies

$$\Lambda_w = (1 - T_w)^2 U_{cc} + 2N_w(1 - T_w) U_{cE} - T_{ww} U_c < 0$$

(b) $\mu = 0$ implies

$$\Lambda_w = U_{cc} + 2(1 - \gamma) U_{cE} + \gamma^2 N_w^2 U_{GG} < 0$$

Shape of Indifference Curves in Reaction Function Diagram with a Constant Tax

$$u = U(w - \tau,\ G,\ N(w) + G) \quad \text{and } \mu = 0 \Rightarrow$$

$$\left.\frac{dG}{dw}\right|_{du=0} = - \frac{U_c + N_w U_E}{U_G + U_E} \quad .$$

Therefore

$$\frac{d}{dw}\left|\frac{dG}{dw}\right._{du=0} = - \frac{1}{(U_G + U_E)} \left[U_{cc} + 2\,\frac{(N_w U_G - U_c)}{(U_G + U_E)}\, U_{cE} + \right.$$

$$\left. + \left(\left.\frac{dG}{dw}\right|_{du=0}\right)^2 U_{GG} \right] > 0$$

REFERENCES

Bruno, M., 1981, "Raw Materials, Profits and the Productivity Slowdown", Working Paper No. 660, *National Bureau of Economic Research.*

Buiter, W., 1980, "The Macroeconomics of Dr Pangloss; A Critical Survey of the New Classical Macroeconomics", *Economic Journal,* March.

Calmfors, L., 1978, "Real Wages, Inflation and Unemployment in the Open Economy", in A. Lindbeck, (ed). *Inflation and Employment in Open Economies,* North-Holland.

Calmfors, L., 1982a, "Cost Adjustment in Smaller OECD Countries", *Skandinaviska Enskilda Banken Quarterly Review,* No. 4.

Calmfors, L., 1982b, "Employment Policies, Wage Formation and Trade Union Behavior in a Small Open Economy", *Scandinavian Journal of Economics,* No. 2.

Faxén, K., 1957, "Monetary and Fiscal Policy Under Uncertainty, Stockholm Economic Studies, New Series 1, Stockholm.

Fellner, W., 1980, "The Valid Core of Rationality Hypotheses in the Theory of Expectations", *Journal of Money, Credit and Banking,* vol. 12, No. 4.

Goldfeld, S., 1982, "Rules, Discretion and Reality", *American Economic Review,* May.

Gordon, R., 1982, "Why US Wage and Employment Behaviour Differs from that in Britain and Japan", *The Economic Journal,* March.

Grossman, G., 1983, "Union Wages, Temporary Lay-Offs and Seniority", *American Economic Review,* June (forthcoming).

Grossman, S., and O. Hart, 1981, "Implicit Contracts, Moral Hazard, and Unemployment", *American Economic Review*, May.

Gylfason, T., and A. Lindbeck, 1982, "Macroeconomic Consequences of Endogenous Governments and Labor Unions", Seminar Paper No. 232, Institute for International Economic Studies, Stockholm.

Helpman, E., 1977, "Nontraded Goods and Macroeconomic Policy Under a Fixed Exchange Rate, *Quarterly Journal of Economics*, August.

Johansen, L., 1977, *Lectures on Macroeconomic Planning, part 1: General Aspects*, North-Holland.

Kouri, P., 1982, "Macroeconomics of Stagflation under Flexible Exchange Rates", *American Economic Review*, May.

Kydland, F., and E. Prescott, 1977, "Rules Rather than Discretion: The Inconsistency of Optimal Plans, *Journal of Political Economy*, June.

Layard, R., 1982, "Is Incomes Policy the Answer to Unemployment", *Economica*, August.

Malinvaud, E., 1982, "Wages and Unemployment", *The Economic Journal*, March.

MacDonald, I., and R. Solow, 1981, "Wage Bargaining and Employment", *American Economic Review*, December.

Negishi, T., 1979, *Microeconomic Foundations of Keynesian Macroeconomics*, North-Holland.

Nickell, S., 1982, "Wages and Unemployment: A General Framework", *The Economic Journal*, March.

Oswald, A., 1979, "Wage Determination in an Economy with Many Trade Unions", *Oxford Economic Papers,* No. 3.

Oswald, A., 1982a, "The Microeconomic Theory of the Trade Union", *The Economic Journal,* September.

Oswald, A., 1982b, "Notes on the Pure Theory of the Trade Union", Institute of Economics and Statistics, Oxford University (mimeo).

Phelps, E., 1978, "Commodity-Supply Shocks and Full-Employment Monetary Policy, *Journal of Money, Credit and Banking,* May.

Sachs, J., 1979, "Wages, Profits, and Macroeconomic Adjustment: A Comparative Study", *Brookings Papers on Economic Activity,* 2.

Sachs, J., 1980, " The Changing Cyclical Behavior of Wages and Prices: 1890-1976, *American Economic Review,* March.

Sachs, J., 1982, "Stabilization Policies in the World Economy: Scope and Skepticism", *American Economic Review,* May.

Söderström, H., and S. Viotti, 1978, "Money Wage Disturbances and the Endogeneity of the Public Sector in an Open Economy", in A. Lindbeck (ed), *Inflation and Employment in Open Economies,* North-Holland.

Towards Full Employment and Price Stability: A Report to the OECD by a Group of Independent Experts (The McCracken Report). 1977.