Trade Liberalisation and Specialisation Within and Across Industries

by Jang Ping Thia

A thesis submitted in January 2008 in fulfilment of the requirements for the degree of Doctor of Philosophy in Economics at the London School of Economics and Political Science

Declaration

I certify that this thesis presented by me in January 2008 for examination for the PhD degree of the University of London is solely my own work.

Signature ....

Date 22 May 2008

Abstract

This thesis investigates three aspects of trade liberalisation. Chapter Two presents a model with business cycle uncertainty, monopolistic competition, and productively heterogeneous firms. The results show that greater trade liberalisation does not always lead to increased firm-level aggregate productivity, since weaker firms can export in the face of adverse home shocks. However, trade liberalisation dampens price-output fluctuations, and is welfare improving if countries have trade partners with uncorrelated shocks. This is a pro-globalisation result since it implies greater macroeconomic stability. Some empirical evidence is presented to support this view.

Chapter Three introduces firm heterogeneity into an Economic Geography setting. The results show that even a small difference in the productivity distributions between two locations can have a significant impact on capital distribution - even as wage-rental rates remain the same across locations - if trade is free enough. It provides an alternative perspective to the Lucas Paradox. The model also shows why high sunk cost industries will locate in less risky locations (North) with greater trade liberalisation, while low sunk cost industries go the other way. Trade liberalisation accentuates these effects, and leads to a different North-South industrial specialisation.

Chapter Four introduces worker skills heterogeneity into an Economic Geography setting. Trade liberalisation occurs in two separate waves. Manufacturing first agglomerates when goods trade is liberalised. The result shows that subsequent services trade liberalisation can lead to a loss in manufacturing (or de-industrialisation), changes in specialisation, and stagnation of manufactur-

ing wages. As a consequence of trade liberalisation, there is inequality both within and between nations. The results also show that a relative increase in skilled workers may lead to greater (not less) skilled workers' premium if it encourages greater services agglomeration. The model is consistent with North-South development patterns.

Contents

1 **Main Introduction** **13**

2 **The Impact of Trade on Aggregate Productivity and Welfare with Heterogeneous Firms and Business Cycle Uncertainty** **17**

2.1 Introduction 17

2.2 The Model Setup 23

2.2.1 Endowments 23

2.2.2 States of the World 23

2.2.3 Preferences 24

2.2.4 Technology and Firms 24

2.2.5 Capital Goods 25

2.3 Equilibrium in Autarky 26

2.3.1 Consumer's Problem 26

2.3.2 Profit Conditions 28

2.3.3 The Impact of Uncertainty and Shocks 30

2.3.4 Aggregate Resource Constraints 31

2.3.5 Equilibrium Characterisation 33

2.4 Opening to Trade 41

2.4.1 Two Country Model 42

2.5 Open Economy with Trade Costs 43

2.5.1 Iceberg Trade Cost 43

2.5.2 Equilibrium Characterisation With Iceberg Cost 44

2.5.3 The Diversification Effect with Free Trade 46

2.5.4 Why Iceberg Trade Cost Matters: Comparison with Melitz Model 50
2.5.5 Extension to Output and Multiple Countries 51
2.6 Empirical Evidence 52
2.6.1 Measures of Trade Diversification 52
2.6.2 Dataset 53
2.6.3 Regression Results 54
2.7 Conclusion 56
2.8 Appendix A 58
2.8.1 Distribution of Aggregate Revenues and Profits 58
2.8.2 Equilibrium Characterisation with Fixed Export Cost . 60
2.8.3 Country Sample 63

3 Why Capital Does Not Migrate to the South: A New Economic Geography Perspective 64

3.1 Introduction 64
3.2 The Model Setup 71
3.2.1 Endowments and Regions 71
3.2.2 Preferences 72
3.2.3 Technology and Firms 72
3.2.4 Pareto Productivity Distributions 76
3.2.5 Sunk Cost 77
3.2.6 Trade Cost 77
3.3 Trade Equilibrium Conditions 78
3.3.1 Export Partitioning 78

3.3.2 Average Profits 79
3.3.3 Productivities of Northern and Southern Firms 80
3.3.4 Aggregate Productivity and Prices 81
3.3.5 Equalisation of Expected Values of Entry in North and South 82
3.3.6 Market Clearing 83
3.3.7 Aggregate Expenditure 85
3.3.8 Equilibrium Solution 85
3.3.9 Solving for Global Manufacturing Labour 86
3.3.10 Solving for Rental-Wage Ratio 87
3.3.11 Solving for Total Composite Resource 87
3.4 When the North Is More Productive 88
3.4.1 Solving for Cutoffs 89
3.4.2 Aggregate and Firm-Level Variables 89
3.4.3 A Numerical Example 90
3.5 The Impact of Risk 93
3.6 Extension to Multi-Industry and Larger North 98
3.6.1 Tomahawk Diagram 99
3.7 Conclusion 100
3.8 Appendix B 103
3.8.1 Calibration of Numerical Simulation 103
3.8.2 Deriving Total Resource Cost 104
3.8.3 Characterising the Break Point 105
3.8.4 Equilibrium Conditions with Mean Preserving Spread for South 106

4 **Evolution of Locations, Specialisation and Returns to Factors With Two Waves of Globalisation** **111**

4.1 Introduction 111

4.2 Model Setup 120

4.2.1 Regions 120

4.2.2 A Hierarchy of Skills and Industries 120

4.2.3 Consumer 123

4.2.4 Production Technology 124

4.2.5 Trade Cost 125

4.2.6 Unit Cost and Prices 126

4.2.7 Profits 126

4.2.8 Workers' Types and Equilibrium Wage Rates 128

4.2.9 Definition of Intermediate Revenues and Total Wage . . 130

4.3 Solving for Equilibrium 131

4.3.1 Equilibrium Conditions 131

4.3.2 Adjustment Process 132

4.3.3 Model Parameters 133

4.4 Results 134

4.4.1 Autarky Results 134

4.4.2 First Wave of Globalisation 134

4.4.3 Second Wave of Globalisation 137

4.5 Factor Responses and Changes in Endowments 142

4.5.1 Comparisons with Neoclassical Theories 143

4.5.2 Offshoring and Wages 146

4.5.3 Worsening Inequality in the North 147

4.5.4 Wage Inequality in the South 148
4.6 Conclusion 148
4.7 Appendix C 151
4.7.1 Patterns of Specialisation in Major Developed and Developing Economies 151
4.7.2 Model Parameters 152
5 Final Conclusion **154**

List of Figures

1 Profit Conditions In Autarky 40

2 Profit Conditions with Free Trade 47

3 Effects on Expected Revenue with Different North-South Productivity Distributions 96

4 Effects of Cutoffs on Expected Revenue (Truncation) 97

5 Tomahawk Diagram with Industries of Two Different Sunk Costs 100

6 Effects of Mean-Preserving Spread of Productivity Distribution 107

7 Share of World Manufacturing 1750 - 1900 112

8 UK Current Account Breakdown into Goods and Services . . . 113

9 Tomahawk Diagram (First Wave) 136

10 North's Equilibrium Wages (First Wave) 138

11 Tomahawk Diagram (Second Wave) North 139

12 Equilibrium Wages (Second Wave) North 141

13 Tomahawk Diagram (Second Wave) with K Type Increase in North 144

14 Equilibrium North Wages with Increase in K 145

15 Skilled and Semi-Skilled Wages in South (Second Wave) 149

List of Tables

1 Parameters for Firms and Business Cycles 40

2 Equilibrium of Business Cycles 40

3 Regression of Growth Standard Deviations 1995-2002 . . . 55

4 Country Samples 63

5 Share of Firms and Capital in More Productive North . . 91

6 Equilibrium Variables with More Productive North 92

7 Share of Firms and Capital in Less Risky North with Different Sunk Cost 94

8 Equilibrium Variables with Riskier South 94

9 Parameters and References 103

10 Current Account Balances (US Millions) of Goods (Top) and Services (Bottom) of Major OECD and Emerging Economies 151

11 Baseline Parameters for Numerical Solutions 153

Acknowledgements

First of all, I would like to express my heartfelt gratitude to my supervisor, Dr Stephen Redding. I first met Steve when he was the lecturer for International Economics and I was an undergraduate at the LSE back in 1998. As my supervisor between 2005 to 2007, he provided invaluable support and tutelage towards the completion of this thesis.

Secondly, I would like to thank Professor Gianmarco I.P. Ottavanio. I had the good fortune to be in Gianmarco's class when he visited LSE in 2006. It was there that I became keenly interested in Economic Geography issues. I would also like to thank faculty members at the LSE and CEP - including Professor Anthony J. Venables, Dr Frederic Robert-Nicoud, Dr Henry Overman, Dr Daniel Sturm, Dr Emanuel Ornelas and Dr Ralph Ossa - for their advice on various parts of the thesis.

Last but not least, I would like to thank my wife Grace for her constant encouragement. My firstborn child Elena also arrived in the spring of 2007, making the years in London fruitful in more ways than one.

The PhD scholarship from Singapore's Ministry of Trade and Industry is also gratefully acknowledged here.

1 Main Introduction

"It has been said that arguing against globalisation is like arguing against gravity" - Kofi Anan, United Nations Secretary General, 1997 - 2007. Except that we probably know and agree a lot more about gravity than globalisation.

The greater integration of goods and factor markets has been the source of keen positive and normative academic debate, and also in terms of theories and empirics. Traditional theories suggest that freer trade must be positive for countries' welfare. For almost two hundred years now, David Ricardo's theory of comparative advantage has stood as an undisputed piece of economic logic. In the 1930s, Eli Heckscher and Bertil Ohlin formalised the idea of comparative advantage through relative endowment differences. In these theories, countries that open themselves to trade will never lose. For many of the years that followed, the positive and normative aspect of analysis almost always favoured greater trade openness[1].

But the international trade narrative has become much more nuanced since the 1980s. Pioneering works that dealt with product differentiation and increasing returns to scale, New Economic Geography, firm-worker heterogeneity, industrial organisations and multi-national firms, political economy, distribution and inequality amongst many others, have provided a wider, deeper, and ultimately richer understanding of the process we call globalisation. Researchers have come to understand how some segments of society, or indeed even some nations, can lose with greater globalisation. Economic Geography models for example, have shown how regions or countries can experience a loss

[1] One notable early exception was by David M.G. Newbery and Joseph E. Stiglitz (1984), where the authors show how pareto inferior trade is possible.

of industries (becoming the periphery) with greater trade liberalisation. The gains of globalisation are also likely to be distributed unevenly. As Professor Venables puts it, "Some countries will experience rapid growth, while others will be left behind."

Today, there is no longer a single globalisation narrative - positive or normative. This may well be more reflective of the rather complicated and multifaceted process that it actually is. Rather than one grand unified theory of international trade, there are now many different models that shine light into different aspects of globalisation. It is in this spirit that this thesis proceeds. Although the thesis broadly investigates the effects of trade liberalisation and specialisation within and across industries, each chapter is self contained in terms of logical flow, equations and notations.

In Chapter Two the thesis attempts to answer an old question: does globalisation make an economy more or less volatile? The traditional theory suggests that greater trade integration makes economies more susceptible to terms of trade shocks. However, the chapter considers an economic environment not with single good sectors, but one with differentiated goods. With business cycle shocks, this can lead to aggregate price-output fluctuations. The surprising result here is that opening to trade may not always increase aggregate productivity. For example, a weak firm that would otherwise have quit the market in the face of an adverse domestic demand shock can continue production if it can export to a high demand market. There is nevertheless a pro-globalisation result. It turns out that even without guaranteeing that aggregate productivity will always increase, opening to trade with another country that has uncorrelated business cycle shocks can dampen domestic price-output fluctuations, thereby providing another source of welfare gains[2].

In Chapter Three, the thesis asks another keenly debated question: why has so little capital moved from the capital-rich North to the capital-scarce South? Furthermore, why has the distribution of investment flow to the South been so uneven? Instead of using a neoclassical aggregate production function as the starting point of the analysis, the paper works with firm heterogeneity. A small perturbation of the ex-ante productivity distribution, from which each firm takes a productivity draw specific to itself, is enough to result in big differences in capital concentration. This is achieved without any wage-rental differences between locations and it therefore provides an alternative resolution to the Lucas Paradox. Trade liberalisation further accentuates the spatial unevenness of the distribution of capital. Another surprising result is that if one location (South) is somehow riskier, without necessarily being less productive on average, high sunk-cost industries will move to the less risky North, while low sunk-cost industries move the other way as trade is liberalised. This leads to both locations specialising in different kinds of industries.

Finally Chapter Four uses an Economic Geography model to analyse the evolution of industrial locations and wages through two waves of globalisation. Goods trade is liberalised in the first wave of globalisation, while services trade is liberalised in the second wave. This is consistent with historical evidence. Shipping costs (for goods) have decreased sharply since the last century, while the fall in communication costs (which pertains to services) is a more recent phenomenon. Using an input-output structure, the model shows how manufac-

[2]The standard 'new' trade model [Paul R. Krugman (1979, 1980)] highlights only the welfare gain through the expansion of varieties.

turing activities can agglomerate in one location during the first globalisation wave. The model then shows that services agglomeration can displace manufacturing in the second globalisation wave when services trade is liberalised. In doing so, the paper touches on the big debate in recent international trade literature - how globalisation might lead to the de-industrialisation of some developed economies, which in turn leads to blue-collar wage stagnation and greater income inequality within nations.

2 The Impact of Trade on Aggregate Productivity and Welfare with Heterogeneous Firms and Business Cycle Uncertainty

2.1 Introduction

Business Cycles and Firm Heterogeneity The business cycle, which exerts a profound impact on many facets of the economy, has generally not been given much consideration in international trade models. Traditional trade theories highlight the gains from trade that arise from country level differences, either broadly due to technology (Ricardian) or endowment (Heckscher-Ohlin). Since these models describe the long run general equilibrium gains from trade, the business cycle is in some sense irrelevant. Perfect competition also renders the firm irrelevant in equilibrium trade considerations. On the other hand, suppose one introduces business cycle shocks to a 'new' trade model [Paul R. Krugman (1979, 1980)] with homogenous firms. Since firms are homogeneous, the business cycle affects all firms symmetrically and does not therefore have any reallocation effects. Any business-cycle driven reallocation of market shares can only be adequately described with a model of heterogeneous firms.

This paper therefore asks the following question: how do business cycle shocks affect heterogeneous firms? What are the reallocative and welfare implications? How do these shocks affect the production and exporting decisions of firms? These questions are interesting and important on several counts.

To begin, trade in the context of firm heterogeneity has received much theoretical research attention recently [Marc J. Melitz (2003); Andrew B. Bernard, Stephen J. Redding and Peter K. Schott (2007) - henceforth known as BRS;

Melitz and Gianmarco I.P. Ottaviano (2005)] motivated by strong empirical evidence that points to the existence of persistent productivity differences between exporters and non-exporters. The key contribution of the firm heterogeneity literature is to formally model the reallocation of output and market shares between productively heterogeneous firms. Firms below the so-called productivity cutoff cease to operate, ceding market shares to more productive firms above the cutoff.

Economists are therefore able to formalise yet another source of welfare gains through trade liberalisation, which arises by increasing the productivity cutoffs and the transfer of market shares to more productive firms. But since these models set out to formalise the long-run equilibrium effects of trade liberalisation, the business cycle is mostly ignored. The notable exception is by Fabio Ghironi and Melitz (2004), which microfounds the Balassa-Samuelson effect through heterogeneous firms and productivity shocks. However, the authors consider only the exporting decisions of firms but not their production decisions.

With heterogeneous firms, it is also evident that business cycle shocks will affect different firms differently, even in autarky. There are potentially interesting reallocative effects of output and market shares. The macro consequence of the business cycle in an environment with heterogeneous firms is non-trivial since a firm's continued production through an adverse demand shock would depend on its productivity. The set of firms that quit production in the face of adverse demand is therefore not a random selection, and there exists a systematic relationship between productivity cutoffs and business cycle shocks[3].

Trade between economies with asymmetric shocks would therefore present another point of interest: how would heterogeneous firms behave and what would be the macroeconomic welfare consequences? The objective of this paper is to model the effect of business cycles and trade in an analytically tractable manner.

Model Outline The starting point of the paper is the introduction of productivity shocks into a Melitz type model. By altering the size of the market, these business cycle productivity shocks then translate into demand shocks for firms. This paper does not consider any nominal rigidities that affect firms' ability to adjust. However, firms have to invest in fixed assets first (due to production lags) before production takes place. The effect of this is that firms face an uncertain demand since they are investing before the shocks are realised[4]. Due to the heterogeneity in production costs, profit outcomes are no longer symmetric. For example, with a negative aggregate shock, weaker firms will make losses while stronger ones will still make profits. With a positive shock, all firms will make profits but again profits will be higher for more productive firms. Though aggregate profits shift up or down

[3] On the other hand, business cycle shocks with homogeneous monopolistically competitive firms do not yield much meaningful analysis. For example, suppose a Krugman type firm has to decide on market entry by making a fixed asset investment without knowing the level of demand entry under uncertainty will occur until ex-ante profit becomes zero for all firms. If demand turns out to be high, there will be insufficient entry and all firms will make a profit. Conversely, there will be too many entry firms if demand is low and all firms will be unable to recover fixed costs and thereby make losses. Depending on the realisation of the aggregate demand shock, either all firms make profits or all firms make losses since firms are homogenous. The equilibrium does not provide any richness in describing the reallocation effect that would occur with heterogeneous firms.

[4] In Ghironi and Melitz (2004), the aggregate shock in that model is introduced via firms' uncertainty over their future productivity. As there are no fixed production cost, production decisions are not affected by shocks - only exporting decisions are affected. The departure in this paper is the presence of fixed production cost, which then affects a firm's decision whether to continue through adverse shocks.

depending on the ex-post demand shocks, there is always a 'profit ranking' where a more productive firm always has a higher profit.

Furthermore, in a general equilibrium, productivity shocks also change the available aggregate resources in the market place. As Melitz (2003) notes, " . . all the effects of trade on the distribution of firms are channelled through a second mechanism operating through the domestic factor market where firms compete for a common resource." The first mechanism - namely product market competition - is "not operative . . . due to CES preferences: the price elasticity of demand for any variety does not respond to changes in the number or prices of competing varieties."[5]

A similar mechanism of factor competition is at work in this paper. When faced with a negative aggregate shock, the aggregate savings (of consumers) fall. Since aggregate savings equal the gross investments into firms' fixed costs, fewer firms are able to invest and continue production. The upshot of this is that weaker firms will have to quit the market, fitting the stylised fact that recessions have a greater impact on weaker firms. In this paper however, an additional mechanism is introduced via demand uncertainty: Firms have to invest in fixed asset before demand is realised. The expected market size (in the next period) will change a firm's decision whether to continue in the market.

Trade and Capital Market Integration Away from autarky, two processes of integration occur. The first is capital market integration that

[5]Melitz and Ottaviano (2005) provide a model with quasi-linear preferences with firm heterogeneity that delivers reallocation of market shares through competition in the goods market. However in that model, any changes to income affect only the consumption of the competitive sector and have no impact on the monopolistic sector. The model is therefore less suitable in the context of modelling demand shocks to the monopolistic sector.

allows capital to be shipped between countries. The second is goods market integration that allows consumption goods to be shipped. As this paper is focused on the effects of goods trade only, capital market integration is assumed to be as simple as possible. There is a perfectly competitive international market for capital to be shipped between countries and returns to capital costlessly remitted back to capital owners for consumption. This is the key assumption of the 'Footloose Capital' class of models in Economic Geography. While this may not necessarily be a robust or realistic assumption, it nevertheless allows the paper to abstract from any capital market complications that might arise and focus on the goods market instead.

It turns out that in equilibrium, even the perfect mobility of capital cannot replicate the outcome of free goods trade[6]. Why might this be so? The presence of trade costs alter the perceived expected market size faced by monopolistically competitive firms. In the presence of trade costs, the productivity cutoffs of two countries cannot be equalised in some circumstances, leading to different selection effects in both countries. The fact that two economies have different productivity cutoffs is not trivial. First, it implies that capital is not optimally invested as some less efficient firms (in the country with lower productivity cutoff) can continue production when they otherwise cannot with free trade. Secondly, it implies that the reallocative effect is not maximised since some weaker firms continue producing behind trade barriers. In the presence of trade costs, market shares are therefore not allocated in the most efficient manner across economies.

[6] This is a different result from Robert Mundell (1957) that shows that free trade in factors is equivalent to free trade in goods in a neoclassical setting.

What then are the benefits of free trade? As firms in each country operate in a larger fully integrated market, the productivity cutoffs in both countries are completely equalised (but may still change with different demand states). This represents the optimum deployment of capital and allocation of market shares between heterogeneous firms and across economies. More significantly, this leads to a diversification effect which in equilibrium reduces price-output fluctuations faced by each economy in autarky. This result stems from the fact that only free trade equalises the productivity cutoffs between countries, and allows for an equally productive set of producers to operate. The increased macroeconomic stability presents yet another source of welfare gains for the risk-averse consumer.

Limitations In this paper, business cycle shocks are introduced by way of a two-state (good or bad) Markov process. While this is more limiting compared to where productivity shocks (innovations) are normally distributed with mean zero [see Ghironi and Melitz (2004)][7], the Markov process allows the paper to solve various variables in a stationary equilibrium. When analysing the effects of international trade, this paper considers only a two-country setting to highlight the diversification effect clearly. Nevertheless, the insights can be extended to a multi-country setup. Finally, to derive analytical solutions explicitly, the paper assumes the productivity distributions to be Pareto [Ghironi and Melitz (2004); BRS (2007)].

[7]For example, it will not be possible to generate variable moments to fit the data, greatly reducing the testable implications on parameters.

2.2 The Model Setup

2.2.1 Endowments

There are L identical consumers (who are also workers) in the economy. The consumers have infinite lives, and each is endowed with some mean level of human capital denoted by H, thereby providing a mean level of effective labour force of LH.

2.2.2 States of the World

There are two possible states of the world, bad and good, denoted by subscript $S = B$ or G. There is high H_G in the good state and low H_B in the bad state. This is the characterisation of the aggregate shock. The transition from period to period is given by a simple Markov process

$$\Pr(H_{G,t+1} \mid H_{G,t}) = \Pr(H_{B,t+1} \mid H_{B,t}) = \rho$$

where $1 > \rho > \frac{1}{2}$ reflects the persistence of the shocks. To abstract from growth dynamics, the model assumes the shocks to be symmetric around the mean level

$$H_G = (1+\gamma)H \qquad\qquad H_B = (1-\gamma)H$$

where $\gamma < 1$ is the size of the shock. Naturally, the average size of this economy over many periods will be LH[8]. Workers sell their labour services to the market inelastically.

[8]There is no long-run growth, and the model abstracts from growth effects considered in Richard E. Baldwin and Frederic Robert-Nicoud (2006).

2.2.3 Preferences

In each period t, the j consumer's utility is given by

$$u_{jt} = x_0^{1-\lambda} x_1^{\lambda} \qquad \text{where } x_1 = \left[\int_{\Omega_t} c_{it}^{\frac{\sigma-1}{\sigma}} di \right]^{\frac{\sigma}{\sigma-1}}$$

and $\sigma > 1$. Good x_0 is the homogenous good, produced competitively with unit labour, costlessly traded, and used as the numeraire ($P_0 = 1$). Good x_1 is the differentiated good, where Ω_t is the set of varieties available to the consumers at discrete period t. Furthermore, each consumer's discounted lifetime utility is given by

$$U_j = \sum_{t=0}^{\infty} \beta^t \ln u_{jt}$$

where $\beta < 1$ is the subjective discount factor in each period.

This preference specification thus exhibits the 'double-diminishing' property. There is diminishing marginal utility to the consumption of each variety in any time period and also diminishing marginal utility to the number of varieties in each period. The log utility also implies that the consumers are strictly risk-averse, preferring a stable level of utility (or varieties) over time. This will be the key property that gives rise to welfare gains when countries trade since aggregate price (or output) stability is welfare enhancing.

2.2.4 Technology and Firms

The homogeneous good is competitively produced. Even after the opening of economies to trade, this paper assumes that the homogeneous good will be produced everywhere (incomplete specialisation), thereby pinning down the price of homogeneous good and wages everywhere to $w = P_0 = 1$ [see Elhanan

Helpman, Melitz and Stephen R. Yeaple (2004)].

For the differentiated industry, there is an exogenous mass M of existing firms with heterogeneous productivity characterised with a productivity distribution that has a cumulative distribution $G(\varphi)$ and density function $g(\varphi)$. Each atomistic firm has a constant productivity φ specific to itself on this distribution[9].

In every period, each firm has a per period fixed cost f. The key requirement is that f has to be in place one period before production takes place due to production lags. If a firm does not invest in f during this period, it will not be able to produce in the period after that. After the fixed cost is incurred, a firm can begin production in the next period with the production function given as

$$l = \frac{q}{\varphi}$$

where l is the labour requirement to produce q units of output.

2.2.5 Capital Goods

Consumers save by investing in a perfectly competitive mutual fund, which then supplies f to the firms in return for next period's operating profits as dividends to the fund. The fund then channels the dividends back to the consumers. This approach is seen in Ghironi and Melitz (2004) and it greatly simplifies the saving-investment process of consumers. With heterogeneous firms, each existing firm will have a different firm value. Considering the investment into a mutual fund this way allows one to ignore the investment

[9] The minimum support of the pareto distribution is given as $\bar{\varphi}$, while the shape is given by parameter k.

choices of individual consumers. This simplification means that the consumers effectively own the entire portfolio of heterogeneous firms through the mutual fund (in equal shares), and receive the same stream of dividend. In this way, one can also characterise the economy with a representative consumer who owns all the firms in the economy.

As countries move away from autarky, this paper makes a departure from Ghironi and Melitz (2004), which assumes that consumers in each country invest in a mutual fund that owns only the portfolio of domestic firms. Here, the paper allows a country's savings to be invested into the fixed cost of firms in another country, and the operating profits to be costlessly remitted back to owners for consumption. This is in a sense the assumption of perfect capital mobility widely used in Footloose Capital models in New Economic Geography.

2.3 Equilibrium in Autarky

As the paper deals with a Markov type uncertainty with only two states, the equilibrium is in fact stationary - the economy switches between good or bad state equilibrium instantly once the shocks are realised, and there are no further transitional dynamics.

2.3.1 Consumer's Problem

Since all consumers are identical, one can deal with the model with a representative consumer (normalising L to 1). The consumer faces a decision on how much to spend (and save) in each period given the state of the world and how much to spend on each good. In each period, the consumer simply solves the following Bellman equation with value function χ

$$\chi_t(\omega_t, H_t) = \max \ln u_t + \beta\chi_{t+1}(\omega_{t+1}, H_{t+1}|S_t)$$

subject to the inter-temporal budget constraint

$$E\left[\frac{\omega_{t+1}}{\iota_t}|S_t\right] = H_t + \omega_t - E_t$$

The representative consumer holds the entire market portfolio of shares of all firms. The first source of income is wage income H_t. The second is the net revenue of firms ω_t returned to the consumer as a dividend. His expenditure is E_t and he saves by again investing in the market portfolio of firms with an expected return ω_{t+1}, suitably discounted by interest rate ι_t.

The optimisation of the Bellman equation with the log utility and Markov process give the following Euler equations

$$\frac{1}{E_G} = \beta\iota_G\left[\rho\frac{1}{E_G} + (1-\rho)\frac{1}{E_B}\right] \qquad \frac{1}{E_B} = \beta\iota_B\left[\rho\frac{1}{E_B} + (1-\rho)\frac{1}{E_G}\right] \quad (1)$$

where E_G and E_B are the expenditures of the good and bad states, ι_G and ι_B are the real interest rates. From equation (1), as expenditure is higher in the good state $E_G > E_B$, the real interest rate is also higher in the bad state $\iota_B > \iota_G$ (the real interest rates will be solved explicitly in later sections). This is a standard result - a higher real interest rate is necessary in the bad state for the consumer to be indifferent between current and future consumption. Since there are only two levels of aggregate expenditure, of which a constant λ is spent on the differentiated sector, there are also only two levels of aggregate revenue in the differentiated sector given by

$$R_G = \lambda E_G \qquad\qquad R_B = \lambda E_B$$

Indirect Utility The indirect utility of the consumer in each period can be written as

$$V = \frac{\lambda^{\lambda}(1-\lambda)^{1-\lambda}E_{S(t)}}{\left[P_{S(t)}\right]^{\lambda}} \tag{2}$$

where S denotes the state. The consumer's indirect utility depends on two factors - his current state-contingent expenditure $E_{S(t)}$ and the aggregate price level $P_{S(t)}$ of the differentiated sector[10]. However, since the number of firms that are producing in period t is determined in period $t-1$ given the production lag, the CES aggregate price level $P_{S(t)}$ in fact depends on the investment decisions in the previous period. For example, if today is a good state while the previous state is bad, the indirect utility is in fact given as $V_B^G = \frac{\lambda^{\lambda}(1-\lambda)^{1-\lambda}E_G}{P_B}$, where $P_G < P_B$. Though today's income is high, welfare is lower due to the higher CES aggregate price. V_G^G therefore gives the highest indirect utility and V_B^B the lowest.

2.3.2 Profit Conditions

The productivity cutoff is defined as a productivity level φ^* that allows a firm to break even in expectation with the investment into fixed cost. A firm with this cutoff productivity level is labelled as the marginal firm. Any firm with a productivity level below this cutoff will not invest in fixed assets and not produce in the next period [see Melitz (2003); BRS (2007)].

[10] The price of the homogenous good is normalised to 1, and therefore does not appear in the indirect utility equation.

Proposition 1 *There exists a 'Profit Condition' for the good state PC_G that provides the relationship between the expected profitability of a firm and its productivity φ, when the economy is hit with a positive shock.*

Proof. In equilibrium, the marginal firm with productivity φ^* will have expected revenue $r^e(\varphi^*)$ that recovers investment cost with interest in expectation only. This can be written as

$$r^e(\varphi^*) = \frac{p(\varphi_G^*)^{1-\sigma}}{P_G^{1-\sigma}} [\rho R_G + (1-\rho)R_B] = \sigma \iota_G f \quad (3)$$

where $p(\varphi_G^*)^{1-\sigma} = \frac{\sigma}{\sigma-1}\frac{1}{\varphi_G^*}$ is the CES optimal price, and $[\rho R_G + (1-\rho)R_B]$ is the expected aggregate market size in the next period given the Markov process. Because of the CES function, the ratio of (expected) revenues between two firms with productivity φ and φ^* is given as $\frac{r^e(\varphi)}{r^e(\varphi^*)} = \left(\frac{\varphi}{\varphi^*}\right)^{\sigma-1}$. This allows the expected revenue of any firm with productivity φ to be expressed as $r^e(\varphi) = \left(\frac{\varphi}{\varphi^*}\right)^{\sigma-1} r^e(\varphi^*)$. The expected profit becomes $\pi(\varphi) = \frac{1}{\sigma}\left(\frac{\varphi}{\varphi^*}\right)^{\sigma-1} \sigma\iota_G f - \iota_G f$. This is simplified to be a function of the productivity cutoff only

$$\pi_G(\varphi) = \left[\left(\frac{\varphi}{\varphi^*}\right)^{\sigma-1} - 1\right] \iota_G f \quad (4)$$

■

Proposition 2 *There exists a bad state Profit Condition PC_B that provides the relationship between the expected profitability of a firm and its productivity φ, when the economy is hit with a negative shock.*

Proof. From the previous proposition, the marginal firm condition becomes

$$r^e(\varphi^*) = \frac{p(\varphi_B^*)^{1-\sigma}}{P_B^{1-\sigma}} \left[\rho R_B + (1-\rho)R_G\right] = \sigma \iota_B f \tag{5}$$

The expected profit, characterised by PC_B, can be written as the function of the marginal firm with productivity φ^* only. As the real interest rate is now ι_B, PC_B can be written as

$$\pi_B(\varphi) = \left[\left(\frac{\varphi}{\varphi^*}\right)^{\sigma-1} - 1\right] \iota_B f \tag{6}$$

Since $\iota_B > \iota_G$, the cost of capital is higher in the bad state, shifting the profit function downwards. The real interest rates will be solved explicitly in later subsections. ■

2.3.3 The Impact of Uncertainty and Shocks

Before the paper proceeds to provide the analytical solution to the equilibrium, it is useful to highlight several key facts of this equilibrium. Four realisations of ex-post profits can occur even though there are only two levels of average productivity (since there are only two cutoffs φ_G^* and φ_B^*), because firms have to make investment decisions before the shocks are realised. Actual profitability is therefore not only a function of productivity but is also affected by ex-post demand. Measuring productivity using ex-post realisations of profit can therefore be misleading. Because of the lag structure, high profits can be due to a positive demand shock without any change in the underlying productivity of firms.

Secondly, firm level aggregates are now affected by the relevant state. In the good state, firms with productivity levels higher than φ_G^* will invest f to produce in the next period. With a negative shock, the cutoff level increases to φ_B^* as market conditions go from easy to tough. The result is that firms between $G(\varphi_B^*)$ and $G(\varphi_G^*)$ will have negative expected profits if they choose to stay in the market.

Since the parameters are constant, the model in fact has stationary equilibrium properties. The equilibrium shifts to the good state or the bad state without any further dynamics. This allows the relationship between the numbers of firms to be written as

$$M_B = \left[\frac{1 - G(\varphi_B^*)}{1 - G(\varphi_G^*)}\right] M_G \tag{7}$$

When a bad state comes after a good one, a proportion of firms $\left[\frac{1-G(\varphi_B^*)}{1-G(\varphi_G^*)}\right]$ will not invest in f and quit the market. The business cycle therefore introduces a selection effect where only a stronger and smaller subset of firms is productive enough to continue investing through the bad state. The final point to make here is that firms below φ_G^* will never invest since they can never recover the fixed cost.

2.3.4 Aggregate Resource Constraints

The aggregate resource constraint for the good state can be written as

$$H(1+\gamma) + \frac{\lambda E_G}{\sigma} = E_G + M_G f \tag{8}$$

The terms on the left hand side are total wage income $H(1+\gamma)$ where γ is the size of the aggregate shock, and dividend $\frac{\lambda E_G}{\sigma}$ which is the operating profits of firms producing in the current period (they invested f previously)[11]. The left hand side thus represents total income flow to the representative worker. The corresponding expression for the bad state can be written as

$$H(1-\gamma)+\frac{\lambda E_B}{\sigma}=E_B+M_B f \tag{9}$$

This paper has done away with the Melitz exit mechanism by assuming a fixed number of firms M on the distribution $G(\varphi)$ [see Thomas Chaney (2006)]. This allows one to write M_G and M_B explicitly as a function of M and the respective cutoffs only

$$M_G=[1-G(\varphi_G^*)]M \qquad\qquad M_B=[1-G(\varphi_B^*)]M \tag{10}$$

This is consistent with equation (7) provided earlier.

Consider the good state aggregate constraint in equation (8). It can be re-written as

$$H(1+\gamma)+E_G\left(\frac{\lambda-\sigma}{\sigma}\right)=M_G f$$

By writing the equation this way, the left hand side of the equation is simply the aggregate savings (net of expenditure). By making use of equation (10), the mass of firms investing in the good state becomes

$$[1-G(\varphi_G^*)]M=\frac{H(1+\gamma)+E_G\left(\frac{\lambda-\sigma}{\sigma}\right)}{f} \tag{11}$$

[11]See Appendix A:2.8.1 for the distribution of revenues and profits across firms. The proofs show that aggregate operating profits (which flow back to consumers as dividend) are functions of aggregate expenditures only, independent of the number of firms. In other words, the distribution of market shares across firms does not affect the aggregate resource constraints.

Similarly, the mass of firms investing in the bad state becomes

$$[1 - G(\varphi_B^*)]M = \frac{H(1-\gamma) + E_B\left(\frac{\lambda-\sigma}{\sigma}\right)}{f} \tag{12}$$

In short, the mass of firms that can carry on investing is a function of the net available resource saved in the economy in each period divided by the per firm capital requirement. These two equations therefore allow the productivity cutoffs to be pinned down once the aggregate expenditure (and hence savings) in each state is known. Since aggregate savings are smaller in a bad state, the productivity cutoff φ_B^* must be higher.

2.3.5 Equilibrium Characterisation

The equilibrium is a set of variables $\{\varphi_G^*, \varphi_B^*, E_G, E_B, \iota_G, \iota_B\}$ that satisfy the pair of Euler equations in (1), resource constraints (11) and (12), and the marginal firm conditions (3) and (5).

Making use of the two Euler equations in (1), the ratio of expenditures can be written as

$$\frac{E_B}{E_G} = \frac{\iota_G}{\iota_B}\left[\frac{\rho\frac{1}{E_G} + (1-\rho)\frac{1}{E_B}}{\rho\frac{1}{E_B} + (1-\rho)\frac{1}{E_G}}\right]$$

Let $\frac{E_G}{E_B} = \theta$, where $\theta > 1$ is the ratio of good to bad state expenditure (θ will be solved later). The above equation can be written as

$$\frac{1}{\theta} = \frac{\iota_G}{\iota_B}\left[\frac{\rho + (1-\rho)\theta}{\rho\theta + (1-\rho)}\right]$$

This gives the ratio of interest rates as

$$\frac{\iota_B}{\iota_G} = \left[\frac{\rho + (1-\rho)\theta}{\rho\theta + (1-\rho)}\right]\theta \tag{13}$$

which is greater than one.

Dividing equation (5) by (3) gives the following relationship

$$\frac{\frac{p(\varphi_B^*)^{1-\sigma}}{P_B^{1-\sigma}}}{\frac{p(\varphi_G^*)^{1-\sigma}}{P_G^{1-\sigma}}}\left[\frac{\rho R_B + (1-\rho)R_G}{\rho R_G + (1-\rho)R_B}\right] = \frac{\iota_B}{\iota_G}$$

This relationship can be simplified in two steps. Firstly, the definition of aggregate prices - which is a function of firm mass and average productivity - can be substituted into the above equation. Secondly, one can make use of the convenient relationship that arise from the pareto distribution - that the ratio of average to cutoff productivity is a constant[12]. This constant is therefore cancelled out on the left hand side of the above equation. Together, these simplify the relationship to

$$\frac{M_G}{M_B}\left[\frac{\rho + (1-\rho)\theta}{\rho\theta + (1-\rho)}\right] = \frac{\iota_B}{\iota_G}$$

By substituting the ratio of interest rates from equation (13), the ratio of firm mass can be solved as

$$\frac{M_G}{M_B} = \theta \tag{14}$$

[12]With the pareto distribution, $\left(\frac{\tilde{\varphi}}{\varphi^*}\right)^{\sigma-1} = \left[\frac{k}{k+1-\sigma}\right]$.

Dividing equation (11) by (12) gives

$$\frac{M_G}{M_B} = \frac{H(1+\gamma) + \theta E_B\left(\frac{\lambda-\sigma}{\sigma}\right)}{H(1-\gamma) + E_B\left(\frac{\lambda-\sigma}{\sigma}\right)}$$

With the left hand side to be exactly θ from equation (14), one can simplify the above relationship and solve for

$$\theta = \frac{1+\gamma}{1-\gamma} \tag{15}$$

as a function of shock parameter γ only. Therefore, the ratio of expenditures $\frac{E_G}{E_B}$ and ratio of firm mass $\frac{M_G}{M_B}$ are exactly the ratio of productivity shocks θ.

From the bad state Euler equation

$$\frac{1}{E_B} = \beta \iota_B \left[\rho \frac{1}{E_B} + (1-\rho)\frac{1}{E_G}\right]$$

Multiplying across by E_G gives $\theta = \beta \iota_B \left[\rho\theta + (1-\rho)\right]$. This allows the real interest rate to be solved as a function of parameters only

$$\iota_B = \frac{\theta}{\beta\left[\rho\theta + (1-\rho)\right]} \tag{16}$$

Similarly, the good state interest rate can solved as

$$\iota_G = \frac{1}{\beta\left[\rho + (1-\rho)\theta\right]} \tag{17}$$

With the solution to the interest rates, one can solve for E_G and E_B by plugging ι_B and ι_G into the marginal firm equations in equations (5) and (3), and then making use of the firm constraint conditions in equations (9) and

(8). These will provide four equations to solve for the remaining endogenous variables E_G, E_B, M_G, and M_B. Nevertheless, because of the complexity of the equations, this method is algebraically cumbersome.

There is a quicker way to solve for the variables. Suppose that $\gamma = 0$ (no shocks). In equilibrium, there will only be one interest rate since $\iota_B = \iota_G = \frac{1}{\beta}$, there will only be one level of expenditure $E = E_G = E_B$, and one constant firm mass $M = M_G = M_B$. The marginal firm condition from equations (5) and (3) collapse to one single equation

$$\frac{1}{M}\psi E = \frac{1}{\beta} f\sigma$$

where $\psi = \left(\frac{\varphi^*}{\tilde{\varphi}}\right)^{\sigma-1} = \left[\frac{k+1-\sigma}{k}\right]$ simply reflects the nice property of the pareto distribution where the ratio of average to cutoff productivity is a constant. Without aggregate shocks, there is also only one aggregate constraint

$$Mf = H + E\left(\frac{\lambda - \sigma}{\sigma}\right)$$

By making the substitution of Mf into the previous relationship, one can solve for

$$E = \frac{\sigma H}{\lambda\beta\psi - \lambda + \sigma} \tag{18}$$

This is the solution to the expenditure level in the absence of shocks ($\gamma = 0$).

Since any shocks are symmetric around the mean level of H, and that $\frac{E_G}{E_B} = \theta = \frac{1+\gamma}{1-\gamma}$ in equilibrium, the exact level of expenditures in the presence of shocks are simply solved as

$$E_G = \frac{\sigma H(1+\gamma)}{\lambda\beta\psi - \lambda + \sigma} \qquad\qquad E_B = \frac{\sigma H(1-\gamma)}{\lambda\beta\psi - \lambda + \sigma} \tag{19}$$

Note that the levels of expenditures depend on parameters only. Firm level variables such as productivity average or cutoff productivities, or aggregate variables such as interest rates, have no bearing at all on the level of expenditures. Fluctuation in expenditures is purely a result of γ with no other influences. With the solutions to the level of aggregate expenditure, one can easily solve for the mass of firms using the aggregate constraints in equations (9) and (8)

$$M_G = \frac{H(1+\gamma)}{f}\left[\frac{\lambda\beta\psi}{\lambda\beta\psi - \lambda + \sigma}\right] \qquad M_B = \frac{H(1-\gamma)}{f}\left[\frac{\lambda\beta\psi}{\lambda\beta\psi - \lambda + \sigma}\right] \tag{20}$$

Aggregate Prices and Welfare Implication The expression of aggregate price is

$$P = M_S^{\frac{1}{1-\sigma}}\left(\frac{\sigma}{\sigma-1}\frac{1}{\tilde{\varphi}_S}\right)$$

where M_S is the number of producing firms with state $S = G$ or B, and $\tilde{\varphi}_S$ is the average productivity defined as

$$\tilde{\varphi}_S = \left[\frac{1}{1-G(\varphi_S^*)}\int_{\varphi_S^*}^{\infty} \varphi^{\sigma-1} g(\varphi)d\varphi\right]^{\frac{1}{\sigma-1}}$$

With the pareto distribution, the average productivity becomes a function of the cutoff only

$$\tilde{\varphi}_S = \left[\frac{k}{k+1-\sigma}\right]^{\frac{1}{\sigma-1}} \varphi_S^*$$

where k is the parameter that characterises the shape of the distribution.

Using the definition of the aggregate prices, the ratio of bad to good CES prices is given as

$$\frac{P_B}{P_G} = \left(\frac{M_G}{M_B}\right)^{\frac{1}{\sigma-1}} \frac{\varphi_G^*}{\varphi_B^*} \tag{21}$$

Following a bad state (due to the lag structure), there are fewer firms and the effect of this is to increase the CES aggregate price. This effect is seen in the term $\left(\frac{M_G}{M_B}\right)^{\frac{1}{\sigma-1}}$ which is greater than 1. However, the average productivity following a bad state rises since only a smaller subset of firms above φ_B^* survive. With firm heterogeneity, there are fewer firms but they are of higher productivity, thereby resulting in an opposite effect on the aggregate price level. This is seen by the ratio $\frac{\varphi_G^*}{\varphi_B^*}$ which is less than 1. Another way of seeing this is to realise that firm heterogeneity softens the effect of underlying shocks because the firms that stop investing f in a bad state are the least productive ones.

Despite the opposing effects, there is no ambiguity on the price level with the pareto distribution. Using the fact that $M_G = \left(\frac{\bar{\varphi}}{\varphi_G^*}\right)^k M$ and $M_B = \left(\frac{\bar{\varphi}}{\varphi_B^*}\right)^k M$ from equation (10) given the pareto distribution, the productivity cutoffs are explicitly solved as

$$\varphi_G^* = \bar{\varphi}\left(\frac{M}{M_G}\right)^{\frac{1}{k}} \qquad \varphi_B^* = \bar{\varphi}\left(\frac{M}{M_B}\right)^{\frac{1}{k}} \tag{22}$$

Substituting these into equation (21), the aggregate price ratio can be solved as

$$\frac{P_B}{P_G} = \theta^{\frac{k+1-\sigma}{k(\sigma-1)}} \tag{23}$$

which is strictly greater than 1 (in other words $P_B > P_G$). Aggregate CES prices are always counter-cyclical. A good state leads to lower prices while a bad state leads to higher prices, amplifying the effect of the business cycle shocks. The larger the γ shock, the larger the fluctuation in aggregate prices and welfare.

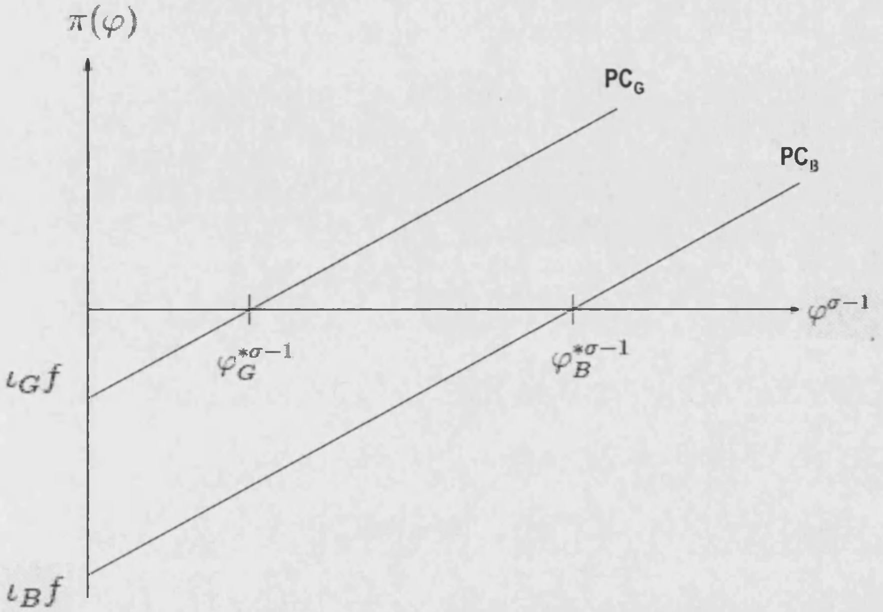
Diagrammatic Representation Diagrammatically, the equilibrium can be illustrated in Figure 1[13]. The profit conditions here are forward looking. Once a firm has invested fixed cost f in the last period, it will definitely produce in the current period because of the CES demands; it does not care about cutoffs. However, the firm has to decide whether to quit or to continue investing f. The Y-axis therefore represents not realised average profits firms earn but expected profits. The X-axis represents the cutoff level of productivity below which firms will choose not to invest in f and quit the market.

Therefore, while there is an exogenous mass of M heterogeneous firms along the entire distribution of $G(\varphi)$, the number of firms that stay in the market is endogenous. Not all are sufficiently productive to stay in the market given the cost of capital. Since aggregate savings are higher in the good state, there will be a larger mass of firms that will invest in f as compared to the bad state $M_G > M_B$. The larger the γ shock, the greater is the firm mass ratio θ, which results in a larger aggregate price fluctuation.

A Simple Numerical Example This subsection provides a simple numerical example to the equilibrium just characterised. The parameters used

[13]Note that by putting φ in the X-axis raised to the power of $\sigma - 1$, the profit conditions become straight lines. The level of capital costs becomes the Y-axis intercepts [see Helpman, Melitz and Yeaple (2004)].

Figure 1: Profit Conditions In Autarky



here are not meant to be realistic as the purpose of this exercise is simply to demonstrate the equilibrium effects in the presence of shocks. The productivity distribution $G(\varphi)$ is assumed to be pareto with support at 0.1 and shape of $k = 4$. The rest of the parameters are provided in Table 1.

Table 1: Parameters for Firms and Business Cycles

Parameter	H	σ	λ	ρ	β	M	f
Value	1000	4	0.5	0.75	0.9	100	1

The equilibrium at two levels of γ shocks are given in Table 2.

Table 2: Equilibrium of Business Cycles

γ	φ_G^*	φ_B^*	ι_G	ι_B	M_G	M_B	E_G	E_B	P_G	P_B
0.05	0.1322	0.1355	1.083	1.138	32.76	29.64	1162.6	1051.8	1.986	2.003
0.10	0.1305	0.1372	1.053	1.164	34.51	28.23	1217.7	996.3	1.977	2.011

fluctuation of the price level between P_B and P_G. Note that the aggregate prices are counter cyclical - a good state leads to lower prices while a bad state leads to higher prices. Given the per period indirect utility in equation (2), the counter-cyclical price fluctuations therefore amplify the effect of expenditures E_G and E_B, resulting in welfare loss for the risk-averse consumer.

2.4 Opening to Trade

Despite firm heterogeneity softening the impact of fewer firms investing in a bad state, there continues to be fluctuation in the aggregate prices caused by business cycle shocks. The important welfare question is: can trade integration between two economies reduce the fluctuation?

In answering this question, a few simplifying assumptions should be made. Firstly, the consumers' expenditures in both economies continue to be uncorrelated after opening to trade. There is no insurance or risk-sharing between consumers of both economies[14]. The implied assumption here is that the international capital market exists for firms only, it does not facilitate borrowing or lending for consumption smoothing. This assumption greatly simplifies the characterisation of the trade equilibrium since it ignores the potential interactions between consumers of two different countries. For the firms, the effect of this assumption is that aggregate demands are uncorrelated across countries. This is not a wholly realistic assumption, but is nevertheless well supported empirically. Indeed, the lack of correlation between consumption of countries is just one of the six major puzzles of international macroeconomics

[14]This could be due to incentives issues such as moral hazard, or costly monitoring and high transaction costs. Because of these reasons, income insurance between countries is not widely observed. Therefore, the trading of international bonds is ruled out.

[see Maurice Obstfeld and Kennedy Rogoff (2000)].

Secondly, there is a perfectly competitive international capital market that allows savings in one economy to be invested towards fixed cost f in another, and net revenue costless remitted back to capital owners for consumption. Consumers (savers) in one economy can invest into and become owners of firms in the other economy in return for next period's profits.

Thirdly, the paper considers only two-country trade for the ease of exposition and to bring out the analytical results more clearly. Nevertheless, as the reader shall see, the insights can be easily extended to multi-country trade.

2.4.1 Two Country Model

Two economies are identical in every way - labour size L, average productivity H, preferences, production technology and productivity distribution. They also have the same mass of firms M on the same productivity distribution $G(\varphi)$. However, both have independent aggregate shocks even after they are open to trade.

Proposition 3 *With free trade, both economies will always have a common productivity cutoff.*

Proof. The proof can be made by contradiction. With free trade, every firm has complete market access into both markets wherever they are located. With free trade, the levels of competitive intensity (characterised by the trade weighted CES price aggregates) are also the same in both locations. Therefore, a firm has to be indifferent between the two locations. Suppose one location (labelled as Home) has a productivity cutoff of φ_H^* while the other (labelled as

Foreign) has a cutoff of φ_F^* such that $\varphi_H^* \neq \varphi_F^*$, a firm that lies between φ_H^* and φ_F^* is above one cutoff (profitable) and below the other cutoff (unprofitable). There exists a mass of firms between φ_H^* and φ_F^* that will not be indifferent since they can invest f in one of the market with positive expected profits. This violates the definition of productivity cutoffs (this proposition will be given a further formal proof later). ■

2.5 Open Economy with Trade Costs

2.5.1 Iceberg Trade Cost

Variable trade costs are introduced as the standard iceberg trading cost of $\tau > 1$ for every unit of good shipped across the economies. With only variable trade cost, the price of export is simply a mark-up over the price of domestic sales $p_X = \tau p$.

Proposition 4 *In the presence of iceberg trade costs, the productivity cutoffs between countries cannot be equalised when they are faced with asymmetric shocks.*

Proof. The paper first sketch a intuitive proof, with the formal proof provided later in the next sub-section. Suppose Home and Foreign economies have asymmetric shocks (Home in a bad state and Foreign in a good state with no loss of generality) and that cutoffs are equalised $\varphi_H^* = \varphi_F^*$. If cutoffs are equalised, the mass of firms investing f is the same in both locations given the assumption of a fixed number (or density) of firms along the same productivity distribution. If the cutoffs are the same at both locations, the aggregate price indices will be equal at both locations whatever the level of trade costs. Since Home is in a bad state, the expected aggregate expenditure, taking into account both domestic and export revenue subjected to trade cost, is

$$\{\rho R_B + (1-\rho)R_G + \phi[\rho R_G + (1-\rho)R_B]\}$$

This is strictly smaller than the expected aggregate expenditure of the Foreign economy

$$\{\phi[\rho R_B + (1-\rho)R_G] + \rho R_G + (1-\rho)R_B\}$$

since it is in a good state and $\phi < 1$ because of trade costs. If φ_H^* defines the firm having zero expected profit if it invests f at Home, a firm with this productivity must have positive expected profits in Foreign given the larger expected market size there. This violates the definition of φ_F^* as the productivity cutoff. ■

2.5.2 Equilibrium Characterisation With Iceberg Cost

This subsection proceeds to characterise the productivity cutoffs in the presence of iceberg trade cost. The impact of fixed export costs f_X will be briefly discussed in Appendix A.

Symmetric Shocks

Proposition 5 *The productivity cutoff, common to both economies, is φ_G^* when they are in the good state; and is φ_B^* when both economies are in the bad state.*

Proof. Consider the case when both economies are in the bad state. Whatever the level of τ, both economies have the expected aggregate revenues since they are hit with symmetric shocks. Furthermore, both economies will have low aggregate savings, with the same aggregate resource constraint in equation (12). Hence, there is no capital flow between the economies. This pins down a common productivity cutoff φ_B^*. The same reasoning applies when both economies are in the good state. ■

Asymmetric Shocks The only case where iceberg cost results in different cutoffs is when Home and Foreign are hit with asymmetric shocks. In this case, aggregate savings in both economies are different and there is the possibility of capital flows affecting the productivity cutoffs in each economy.

Without a loss of generality, suppose Home economy has the bad state while Foreign has the good state, and that trade cost is positive $\tau > 1$. Given $E_H = E_B$ and $E_F = E_G$[15], the trade equilibrium is a set of variables $\{\varphi_H^*, \varphi_F^*, \iota_M\}$ that satisfy the following conditions, where ι_M is the cost of capital faced by the firms in the Home and Foreign economy respectively.

First, the marginal firms with productivities φ_H^* and φ_F^* must have zero profits in their respective locations. This gives the pair of equations

$$\frac{p(\varphi_H^*)^{1-\sigma}}{P_H^{1-\sigma}}[\rho R_B + (1-\rho)R_G] + \frac{\phi p(\varphi_H^*)^{1-\sigma}}{P_F^{1-\sigma}}[\rho R_G + (1-\rho)R_B] = \iota_M \sigma f$$

$$\frac{\phi p(\varphi_F^*)^{1-\sigma}}{P_H^{1-\sigma}}[\rho R_B + (1-\rho)R_G] + \frac{p(\varphi_F^*)^{1-\sigma}}{P_F^{1-\sigma}}[\rho R_G + (1-\rho)R_B] = \iota_M \sigma f$$

where $p(\varphi_H^*) = \frac{\sigma}{\sigma-1}\frac{1}{\varphi_H^*}$ and $p(\varphi_F^*) = \frac{\sigma}{\sigma-1}\frac{1}{\varphi_F^*}$ are the optimal prices charged by the marginal firms, P_H and P_F are the trade weighted CES price aggregates.

[15] Note that from equation (19), since there is no insurance across consumers in the different countries, their levels of expenditures are affected by their domestic shocks only.

By substituting the expressions for the CES price aggregates and cancelling out some terms, the above equations become simplified to

$$\frac{\varphi_H^{*\sigma-1}\left[\rho R_B+(1-\rho)R_G\right]}{M_H\tilde{\varphi}_H^{\sigma-1}+\phi M_F\tilde{\varphi}_F^{\sigma-1}}+\frac{\varphi_H^{*\sigma-1}\phi[\rho R_G+(1-\rho)R_B]}{\phi M_H\tilde{\varphi}_H^{\sigma-1}+M_F\tilde{\varphi}_F^{\sigma-1}}=\iota_M\sigma f \tag{24}$$

$$\frac{\varphi_F^{*\sigma-1}\phi[\rho R_B+(1-\rho)R_G]}{\phi M_H\tilde{\varphi}_H^{\sigma-1}+M_F\tilde{\varphi}_F^{\sigma-1}}+\frac{\varphi_F^{*\sigma-1}\left[\rho R_G+(1-\rho)R_B\right]}{M_H\tilde{\varphi}_H^{\sigma-1}+\phi M_F\tilde{\varphi}_F^{\sigma-1}}=\iota_M\sigma f \tag{25}$$

Secondly, given the global pool of savings which is the resource constraint, the total number of firms is given as

$$[1-G(\varphi_H^*)]M+[1-G(\varphi_F^*)]M=\frac{H(1+\gamma)+E_G\left(\frac{\lambda-\sigma}{\sigma}\right)}{f}+\frac{H(1-\gamma)+E_B\left(\frac{\lambda-\sigma}{\sigma}\right)}{f} \tag{26}$$

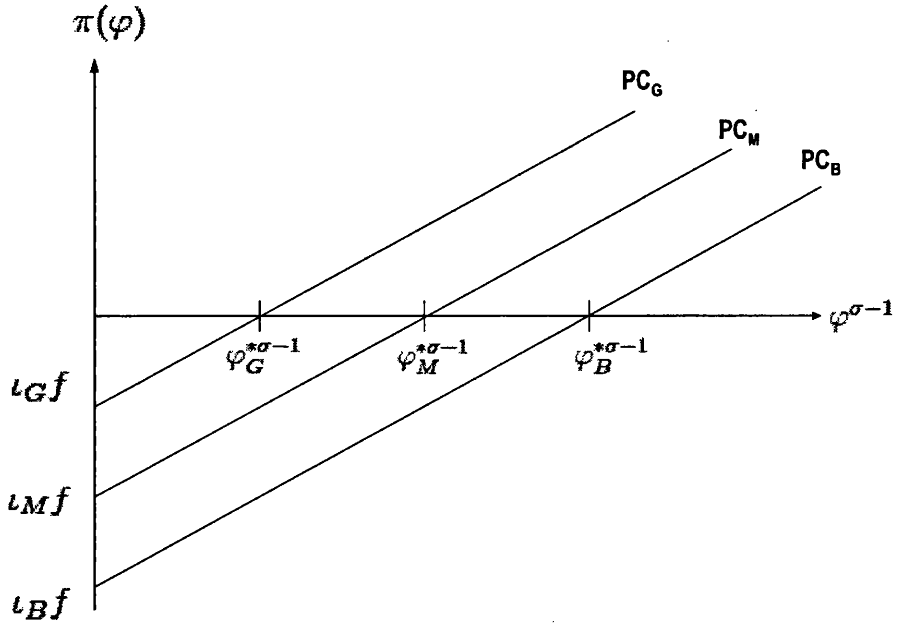
Together, these provide three conditions to solve for $\{\varphi_H^*, \varphi_F^*, \iota_M\}$ given E_G and E_B. Note that it must be the case that $\varphi_H^* \geqslant \varphi_F^*$ in equilibrium since there is a smaller set of producers for the Home country (which is in a bad state)[16].

2.5.3 The Diversification Effect with Free Trade

Given the characterisation of the equilibrium with costly trade (positive iceberg costs), it is easy to show the equilibrium effects under free trade. With

[16] Note that even though trade cost is positive, the two economies continue to have a common cost of capital ι_M because capital is completely mobile. That is, the last unit of capital f invested must recover the same expected amount $\iota_M f$ in both economies even though they have asymmetric shocks. Therefore, $\iota_H = \iota_F = \iota_M$ in equilibrium even in the presence of positive trade cost. The fact that productivity cutoffs are not equalised is due to trade cost altering the degree of capital flows between the two economies when they have asymmetric shocks.

Figure 2: Profit Conditions with Free Trade



equations (24) and (25), one arrives at the following equality

$$
\begin{aligned}
& \frac{\varphi_H^{*\sigma-1}\left[\rho R_B + (1-\rho) R_G\right]}{M_H \tilde{\varphi}_H^{\sigma-1} + \phi M_F \tilde{\varphi}_F^{\sigma-1}} + \frac{\varphi_H^{*\sigma-1}\phi[\rho R_G + (1-\rho) R_B]}{\phi M_H \tilde{\varphi}_H^{\sigma-1} + M_F \tilde{\varphi}_F^{\sigma-1}} \\
= & \frac{\varphi_F^{*\sigma-1}\phi[\rho R_B + (1-\rho) R_G]}{\phi M_H \tilde{\varphi}_H^{\sigma-1} + M_F \tilde{\varphi}_F^{\sigma-1}} + \frac{\varphi_F^{*\sigma-1}\left[\rho R_G + (1-\rho) R_B\right]}{M_H \tilde{\varphi}_H^{\sigma-1} + \phi M_F \tilde{\varphi}_F^{\sigma-1}}
\end{aligned}
$$

With free trade ($\phi = 1$), the aggregate prices - given by the denominators - are equal. The expected revenues in brackets also become equal. Together, these imply that $\varphi_H^* = \varphi_F^*$ even in the presence of asymmetric shocks. A Home firm will be a perfect substitute for the foreign firm. In other words, free trade will result in a common cutoff φ_M^* even with asymmetric shocks to both economies. This is the formal proof to Proposition 3. The effects of free trade can be seen in Figure 2.

This result can also be inferred from the firm mass equations in (20). These give the firm masses in equilibrium with the good and bad state, which is a function of γ shocks and other parameters only. In a fully integrated economy with free trade, it simply means that the γ shocks cancel out. With free trade, the firm mass that is common to both economies becomes

$$M_M^{FT} = \frac{H}{f}\left[\frac{\lambda\beta\psi}{\lambda\beta\psi - \lambda + \sigma}\right] \tag{27}$$

Since there is a common firm mass, there is a common cutoff φ_M^* [analogous to the relationships specified in equation (10)][17].

From equation (23), the ratio of aggregate prices is a function of productivity cutoffs only. Denoting free trade variables with superscript FT, the price ratios become $\frac{P_B^{FT}}{P_G^{FT}} = \theta^{\frac{k+1-\sigma}{k(\sigma-1)}}$, where P_B^{FT} denotes the aggregate price when both economies are faced with a negative shock (analogous definition for P_G^{FT})[18]. With asymmetric shocks, the productivity cutoff becomes φ_M^* for both economies (where $\varphi_B^* > \varphi_M^* > \varphi_G^*$) given free trade. As the γ shocks are cancelled out, the aggregate price level with asymmetric shocks P_M^{FT} therefore lies between P_B^{FT} and P_G^{FT}.

There are now two sources of gains from trade. Firstly there is an expansion of varieties leading to lower aggregate prices and higher welfare. Secondly, there is a reduction in the probability that extreme price levels are reached. This reduces the variance in the aggregate price level and the fluctuation in real

[17] In the presence of trade cost, the firm mass is always larger in the country with the positive shock since the expected market size is bigger. The economy with the negative shock will have a smaller firm mass and higher price aggregate. Without free trade, aggregate prices are not equalised with asymmetric shocks. Given that consumers are risk-averse, this is welfare-reducing.

[18] Note that since both economies are in the same state, the cutoffs are unchanged from the autarky counterparts, which are φ_B^* for the bad state and φ_G^* for the good state.

income, thereby representing a welfare gain from diversification for the risk-averse consumer. This is a gain from trade above and beyond the expansion of variety effect.

Free trade therefore results in the optimal allocation of market shares for there will always be an equally productive subset of firms producing in each economy and selling across markets. The result is that productivity cutoffs are completely equalised even as countries face asymmetric shocks.

When economies are hit with asymmetric shocks, there is essentially a diversification equilibrium. For example, suppose the Home economy is in a bad state. Under autarky, the cutoff productivity would have been φ_B^*. However, if the trading partner Foreign is in a good state, Home's cutoff productivity falls to φ_M^* with free trade. In other words, some firms that would have quit a domestic negative shock at Home in autarky will now continue to produce, as expected profits from exporting more than compensate for the expected domestic loss. Aggregate firm-level productivity therefore does not always increase with trade.

Finally, there is a subtle implication from the price ratio $\frac{P_B^{FT}}{P_G^{FT}} = \theta^{\frac{k+1-\sigma}{k(\sigma-1)}}$. The parameter k characterises the level of firm heterogeneity. A smaller k implies that firms are more heterogenous while a larger k implies that firms are more homogenous. A decrease in k would lead to an increase in $\frac{k+1-\sigma}{k(\sigma-1)}$, propagating the aggregate price fluctuations. In other words, if firms are more heterogeneous, the net effect (after accounting for firm entry and changes in aggregate productivity) is greater price fluctuation. This suggests that the diversification gains from trade are higher when firms are more heterogeneous.

2.5.4 Why Iceberg Trade Cost Matters: Comparison with Melitz Model

In the Melitz model with the absence of fixed export cost, the passage from autarky to free trade (by τ falling from infinity to 1) increases welfare through the CES price aggregates, with no further impact on firm level variables. The reason for this is that a fall in τ increases local competitive intensity through the price index but also increases export revenue, leaving the firm exactly indifferent.

However, the level τ is crucial here and affects the productivity cutoffs. The key here is to realise that Melitz presents a model which is a long run stable equilibrium of countries of symmetric sizes, "Firms correctly anticipate this stable aggregate environment when making all relevant decisions. The analysis then focuses on the long run effects of trade and the relative behaviour and performance of firms with different productivity levels." In that model, both consumption demand and investment into firms are constant. The presence of iceberg cost therefore does not have any further effect since it preserves the homotheticity amongst all firms.

In this paper, even though both economies have a long run average size of LH, each of them fluctuates around two states defined by the Markov process. The pool of aggregate savings in each economy changes according to the shocks, thereby changing the resource available for investment, and in the process altering the survivability conditions in different states.

Cross Border Capital Flow Given that aggregate savings are not the same when the countries are faced with asymmetric shocks, there will be cross border capital flow. Through its effects on expected market potentials, τ changes the incentive for cross-border capital flows. A lower τ provides higher incentive for the high savings economy (good state) to invest into the low savings economy (bad state), until the productivity cutoffs are completely equalised with free trade. Conversely, a higher trade cost τ creates a divergence between the perceived market sizes when the economies are hit with asymmetric shocks and reduces the diversification effect. Trade liberalisation therefore dampens differences in productivity cutoffs between two economies when they are hit with asymmetric shocks, leading to lower price-output fluctuations.

The key point is this: free capital mobility, in the presence of positive trade costs, cannot equalise the productivity cutoffs between two economies when they are hit with asymmetric shocks. Therefore, free capital mobility alone cannot replicate free trade outcomes. Since productivity cutoffs are unequal with asymmetric shocks and positive trade costs, market shares are not allocated in the most efficient way between heterogeneous firms across the two economies. Only with free trade will there be optimal allocation of market shares between productively heterogeneous firms across countries - that is, an equally productive subset of producers in each country (above a common productivity cutoff of φ_M^* when there is asymmetric shocks) will stay in the market.

2.5.5 Extension to Output and Multiple Countries

In the setup of the model, the paper has modelled welfare changes to the consumer through the impact of trade on the CES price aggregates. However, there is an simple conceptual extension to output. If one considers the differentiated sector as an immediate sector supplying a final competitive sector as in the Ethier production function [see also Anthony J. Venables (1996)], the smaller price fluctuation shown here directly translates into smaller output fluctuation of the final sector. Free trade therefore reduces output fluctuation in this interpretation. So long as the consumer is risk-averse, the lower fluctuation of price-output will be a source of welfare gain.

Furthermore, if a large number of countries with uncorrelated γ shocks are engaged in free trade, all of them will converge to φ_M^*, completely stabilising aggregate price-output across all economies. Except to note that this result is obvious from the Central Limit Theorem, this will not be given any formal proof.

2.6 Empirical Evidence

The key result in this paper is that trade between economies with uncorrelated productivity shocks leads to lower price-output fluctuation. This sub-section presents a stylized test to the hypothesis: whether there is indeed a (negative) correlation between trade diversification - as measured by the spread of trade across different trading partners - and output volatility.

2.6.1 Measures of Trade Diversification

The measure of trade diversification is the equivalent of the Herfindahl index. Take exports as an example. For each i country in the sample, the share of export to each of the other countries j in the sample is computed s_{ij}. The

Export Dispersion Index (EDI) is given as

$$EDI_i = \sum_{j \neq i} s_{ij}^2$$

which is the sum of squares of the shares of export. EDI is calculated using 1995 data. A measure of 1 will mean that the country is exporting all its merchandise good to one market - that is, its export is highly concentrated. A measure approaching 0 will imply the greatest possible export market diversification.

Another two dispersion indices are calculated for each country using year 1995 data, the Import Dispersion Index (IDI) and Trade Dispersion Index (TDI), the last being a trade-weighted average of EDI and IDI. TDI therefore fully captures how diversified a country's trade is, or broadly how integrated it is with the rest of the world. The measure of output volatility is taken to be the standard deviation of annual GDP growth between 1995 and 2002.

2.6.2 Dataset

The dataset is from the Direction of Trade Statistics (DOTS) and is compiled by the IMF. It provides the merchandise exports and imports reported by each country with all trading partners[19]. There are a total of 115 countries in the sample [full sample provided in Table 4 in Appendix A:2.9.3]. This is considerably smaller than the original data coverage for several reasons. Firstly, dependencies and territories are generally dropped. Secondly, countries known to be afflicted by armed conflicts, either international or civil, during the pe-

[19] Unfortunately, this dataset does not include services trade and a similar dataset for services is not available. The data therefore does not fully capture the extent of trade diversification.

riod are also dropped. Thirdly, countries under known UN sanctions are also dropped. The motivation for dropping these countries is that their observed growth volatility and trade shares are driven by considerations from which the model abstracts.

2.6.3 Regression Results

Cross-section regressions are then performed with growth volatility as the independent variable, dispersion indices as the dependent variables along with other controls[20]. Five regressions are performed. In regression (1), none of the dispersion indices are included. EDI is added to regression (2), IDI to regression (3) and TDI to regression (4). Naturally, EDI, IDI and TDI exhibit a high degree of correlation, which is why they are used as independent variables separately rather than altogether. Finally in (5), the regression is run using TDI and with log population as the analytical weight. The results are presented in Table 3.

Not altogether surprising, ex-Soviet and crisis-affected countries have higher growth deviations. The EU and NAFTA dummies show lower growth deviations. The coefficient for GDP size is negative (as expected) for all cases. However, it is only significant in regression (1), before the introduction of various dispersion indices. Furthermore, the constant term in regression (1) is high, and highly significant.

Subsequent regressions that include EDI, IDI and TDI all show an im-

[20]Some country group are captured by the various dummies: Crisis, Ex-USSR, OPEC, EU and NAFTA. Crisis countries are those affected by financial crisis during the sample period, and include Thailand, Malaysia, Korea and Argentina. A dummy is also created for ex-Soviet bloc countries, which experienced a difficult transition to free markets during much of the 1990s.

Table 3: Regression of Growth Standard Deviations 1995-2002

	(1)	(2)	(3)	(4)	(5)
Log(GDP)	-0.155*	-0.044	-0.065	-0.041	-0.045
	(0.076)	(0.085)	(0.082)	(0.084)	(0.087)
Goods Trade as % of GDP	0.002	0.001	0.000	-0.000	0.000
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Crisis	3.533**	3.531**	3.395**	3.464**	3.477**
	(0.358)	(0.350)	(0.358)	(0.353)	(0.350)
Ussrbloc	1.839**	1.797**	1.830**	1.844**	1.895**
	(0.570)	(0.534)	(0.544)	(0.539)	(0.552)
Opec	0.569	0.296	0.679	0.438	0.419
	(0.466)	(0.404)	(0.446)	(0.402)	(0.397)
EU	-0.758*	-0.811*	-0.953**	-0.943**	-0.944**
	(0.304)	(0.330)	(0.315)	(0.321)	(0.320)
NAFTA	0.448	-1.754+	-1.057	-1.607+	-1.524
	(0.807)	(1.036)	(0.860)	(0.962)	(0.936)
EDI		4.962**			
		(1.529)			
IDI			4.367*		
			(1.972)		
TDI				5.172**	5.116**
				(1.736)	(1.727)
Constant	6.115**	2.860	3.488	2.812	2.896
	(1.903)	(2.184)	(2.111)	(2.160)	(2.242)
Observations	115	114	113	113	113
R-squared	0.37	0.42	0.40	0.42	0.43

Robust standard errors in parentheses
+ significant at 10%; * significant at 5%; ** significant at 1%

provement of fit, with a reduction in both the size and significance of the constant term. Even the GDP term, which has some explanatory power in regression (1), becomes insignificant when the trade dispersion indices are included. All three trade dispersion indices are significant and improve the overall fit. In particular, including TDI results in the best fit and a highly significant coefficient.

The results contrast with some existing research. For example, Dani Rodrik (1998) shows that countries that trade more tend to have bigger governments. The maintained hypothesis is that greater trade integration has resulted in greater economic volatility, and hence necessitating bigger governments as demand stabilisers. The results here however show that openness (measured by trade as a percentage of GDP) does not have explanatory power over output volatility. On the other hand, the results show that diversifying trade with more trade partners is associated with lower output volatility and hence higher welfare. This can therefore be interpreted as a pro-globalisation result.

Given that countries are by definition heterogeneous, and that the regression setup highly parsimonious, the result has to be interpreted with some caution. The quality of institutions, for example, will affect both the ability of firms to carry out trade with the rest of the world and also the fluctuation of growth rates. On the other hand, firms in countries whose domestic demand may be more prone to shocks could have more incentives to seek out other export markets. The results should therefore be interpreted as a correlation between growth volatility and trade diversification (rather than causation). Nevertheless, the result provides some evidence that trade diversification is indeed correlated with lower aggregate output volatility, and suggests that further research into this issue is warranted.

2.7 Conclusion

This paper has built on recent trade and firm heterogeneity literature, in particular the aggregation properties of Melitz (2003) in the presence of firm heterogeneity. Real Business Cycles type aggregate productivity shocks, with consumers who optimise inter-temporally, are introduced into a setting where firms have to invest in fixed assets before the realisation of the shocks. This model therefore makes the firm's problem more realistic compared to traditional trade models.

Without firm heterogeneity, a negative shock would result in all firms making losses, thereby rendering any between-firms analysis meaningless. As it is now possible to solve for market and firm level outcomes in the presence of firm heterogeneity, it has become meaningful to analyse the reallocative impact of such shocks. Different firms will be affected differently while still allowing for market aggregates to be solved analytically.

When trade is not totally free, the productivity cutoffs cannot be equalised, and some producers are shielded by trade barriers and will continue to have positive market shares. With free trade, productivity cutoffs are always equalised in both economies in a full diversification outcome - which means that an equally productive subset of producers remain in the market. Nevertheless, the diversification equilibrium also implies that aggregate firm-level productivity does not always increase with trade. Weak companies that would have quit in a negative demand shock in autarky can continue to operate given diversification possibilities.

Despite this, the model offers a comforting result for trade economists by identifying another source of trade gains. The key to unlocking the insight from the model lies in understanding that opening to trade results in smaller fluctuation in the aggregate price levels and may therefore raise the welfare of risk-averse consumers.

2.8 Appendix A

2.8.1 Distribution of Aggregate Revenues and Profits

This subsection highlights the distribution of aggregate revenues and profits across firms. It will show that aggregate revenues are independent of the number of existing firms. The stream of dividend to the consumers, which depends on aggregate revenues only, is therefore also unaffected by the number of firms. This shows that the good and bad state resource constraints in equations (8) and (9) are also independent of firm level considerations.

The consumer is forward looking. Once the current state is realised, his adjustment to E_G or E_B is instant, pinning down the current period's market size or aggregate revenue for the industry (R_G or R_B). However, the revenue per firm depends on the number who invested f in the previous period, which depends on the last period's realised state. There could either be M_G or M_B firms investing f previously, who will share revenue this period

$$R_G = M_B \bar{R}_{G,B} = M_G \bar{R}_{G,G}$$

where $\bar{R}_{G,B}$ denotes the average per firm revenue conditioned on a previously bad state (analogous for $\bar{R}_{G,G}$). Since $M_G > M_B$, the per firm revenue is higher when there are fewer competitors $\bar{R}_{G,B} > \bar{R}_{G,G}$. Similarly for the bad state

$$R_B = M_B \bar{R}_{B,B} = M_G \bar{R}_{B,G}$$

where $\bar{R}_{B,B} > \bar{R}_{B,G}$. Therefore, conditioning out the current state, average revenue is always higher when the previous state is bad. Since the ratio of average productivity is directly related to the ratio of average productivity (to the power of $\sigma - 1$), this shows that average productivity of firms is higher following a bad state.

$$\frac{\bar{R}_{G,B}}{\bar{R}_{G,G}} > 1 \Rightarrow \left(\frac{\tilde{\varphi}_B}{\tilde{\varphi}_G}\right)^{\sigma-1} > 1 \Leftarrow \frac{\bar{R}_{B,B}}{\bar{R}_{B,G}} > 1 \tag{A1}$$

This result shows that $\varphi_B^* > \varphi_G^*$. Conditioned on the current state, average profit is therefore also higher if the previous state is bad.

Per firm profit is higher following a bad state To develop this idea more formally, one can show that

$$R_{G,B} = \int_{\varphi_B^*}^{\infty} \frac{p(\varphi)^{1-\sigma} E_G}{P_B^{1-\sigma}} M_B g(\varphi) d\varphi = \int_{\varphi_G^*}^{\infty} \frac{p(\varphi)^{1-\sigma} E_G}{P_G^{1-\sigma}} M_G g(\varphi) d\varphi = R_{G,G}$$

It does not matter what the previous state is, aggregate revenue R_G depends on only the current state. Furthermore, operating profit will also be $\frac{R_G}{\sigma}$ if today is a good state. Similarly

$$R_{B,B} = \int_{\varphi_B^*}^{\infty} \frac{p(\varphi)^{1-\sigma} E_B}{P_B^{1-\sigma}} M_B g(\varphi) d\varphi = \int_{\varphi_G^*}^{\infty} \frac{p(\varphi)^{1-\sigma} E_B}{P_G^{1-\sigma}} M_G g(\varphi) d\varphi = R_{B,G}$$

This establishes the following inequalities

$$\bar{R}_{G,B} = \frac{R_{G,B}}{M_B} > \frac{R_{G,G}}{M_G} = \bar{R}_{G,G} \qquad \bar{R}_{B,B} = \frac{R_{B,B}}{M_B} > \frac{R_{B,G}}{M_G} = \bar{R}_{B,G} \tag{A2}$$

Since the consumers optimise instantly, R_G and R_B are pinned down immediately. However, the number of firms selling in this period has the lag effect of investing f the previous period. Aggregate revenue is therefore shared among the mass of firms determined in the previous period, and the market shares allocated as such. However, aggregate revenues are unaffected by the number of firms since $R_{G,B} = R_{G,G}$ (good state) and $R_{B,B} = R_{B,G}$ (bad state). The stream of dividend for consumers in each state is therefore also unaffected by the number of firms.

2.8.2 Equilibrium Characterisation with Fixed Export Cost

The firm heterogeneity literature is motivated by the empirical evidence that only a small and productive subset of firms engage in exporting activities. The presence of iceberg trade cost alone does not create this export partitioning, due to the CES preferences. In order to achieve export partitioning, a fixed export cost f_X has to be introduced. This paper assumes that f_X has exactly the same conditions attached to f - it is funded through aggregate savings and has to be invested one period before export can take place.

For exporters to be a small and more productive subset of all firms, there must exist firms with productivity below φ that find it profitable to operate domestically (with domestic revenue r_D) but not export (thereby foregoing revenue r_X). The two inequalities therefore become

$$\frac{r_D(\varphi)}{\sigma} - f > 0 \qquad\qquad \frac{r_D(\varphi)}{\sigma} + \frac{r_X(\varphi)}{\sigma} - f - f_X < 0$$

where $r_X(\varphi) = \tau^{1-\sigma} r_D(\varphi)$ due to the CES preference. Together, the partitioning condition implies that

$$f < \tau^{1-\sigma} f_X$$

which says that the combination of iceberg cost and fixed export cost must be high enough to deter some firms from exporting. Define φ_X^* as the marginal firm that just breaks even through exporting. The probability that a firm is strong enough to export is the conditional probability of a firm having a distribution above φ_X^*. This conditional probability $\tilde{p}_X$ is given as

$$\tilde{p}_{HX} = \frac{1 - G(\varphi_{HX}^*)}{1 - G(\varphi_H^*)}$$

Suppose that both economies are in the same state. Note then that $r_D(\varphi_H^*) = \sigma\iota_H f$ and $r_X(\varphi_{HX}^*) = \phi r_D(\varphi_{HX}^*) = \sigma\iota_H f_X$. Taking ratios of the two gives the following relationship

$$\phi\left(\frac{\varphi_{HX}^*}{\varphi_H^*}\right)^{\sigma-1} = \frac{f_X}{f} \qquad \text{or} \qquad \varphi_{HX}^* = \frac{1}{\tau}\left(\frac{f_X}{f}\right)^{\frac{1}{\sigma-1}}\varphi_H^*$$

which says that the export cutoff φ_{HX}^* is a function of domestic cutoff φ_H^* only. This allows the conditional export probability $\tilde{p}_X$ to be determined as a function of parameters only.

Suppose Home and Foreign are in a good state. The aggregate resource constraint from equation (26) becomes modified as

$$[1 - G(\varphi_G^*)]M = \frac{H(1+\gamma) + E_G\left(\frac{\lambda-\sigma}{\sigma}\right)}{f + \tilde{p}_{HX} f_X} \tag{A3}$$

In other words, global aggregate savings have to be used to fund fixed cost f as well as the the f_X requirements of exporters. By inspecting equation (A3) and comparing it with equation (26), it is clear that the effect of fixed export cost will shift φ_G^* rightwards (higher). The effect of f_X creates another source of resource competition. As exporters demand f_X, there will be fewer firms in equilibrium and productivity cutoffs will have to increase. Similar analytical reasoning can be applied to when both economies are in a bad state or when they have asymmetric shocks. The effect of fixed export cost will always push productivity cutoffs higher.

Assuming that export partitioning holds, the conditional probability of exporting $\tilde{p}_{HX}$ is strictly less than 1. The effects of trade liberalisation (as characterised by a fall in τ) can be seen from the above equation. As τ falls, the conditional probability of exporting increases. This increases the denominator of equation (A3), leading to an increase in the productivity cutoffs φ_G^* through the competition of resource.

2.8.3 Country Sample

This is the list of countries.

Table 4: Country Samples

Albania	Egypt, Arab Rep.	Latvia	Slovak Republic
Algeria	El Salvador	Lebanon	Slovenia
Angola	Estonia	Libya	South Africa
Argentina	Ethiopia	Lithuania	Spain
Armenia	Finland	Luxembourg	Sri Lanka
Australia	France	Macao, China	Sudan
Austria	Gabon	Malaysia	Sweden
Azerbaijan	Gambia, The	Mexico	Switzerland
Bahrain	Georgia	Mongolia	Syrian Arab Republic
Bangladesh	Germany	Morocco	Tanzania
Belarus	Ghana	Mozambique	Thailand
Belgium	Greece	Netherlands	Trinidad and Tobago
Bolivia	Guatemala	New Zealand	Tunisia
Brazil	Honduras	Nicaragua	Turkey
Bulgaria	Hong Kong, China	Nigeria	Uganda
Cambodia	Hungary	Norway	Ukraine
Cameroon	Iceland	Oman	United Arab Emirates
Canada	India	Pakistan	United Kingdom
Central African Republic	Indonesia	Panama	United States
Chile	Iran, Islamic Rep.	Paraguay	Uruguay
China	Ireland	Peru	Uzbekistan
Colombia	Israel	Philippines	Venezuela, RB
Costa Rica	Italy	Poland	Vietnam
Cote d'Ivoire	Jamaica	Portugal	Yemen, Rep.
Croatia	Japan	Romania	Zimbabwe
Czech Republic	Jordan	Russian Federation	
Denmark	Kazakhstan	Saudi Arabia	
Dominica	Kenya	Senegal	
Dominican Republic	Korea, Rep.	Serbia and Montenegro	
Ecuador	Kuwait	Singapore	

3 Why Capital Does Not Migrate to the South: A New Economic Geography Perspective

3.1 Introduction

It has long been a source of consternation among economists as to why there has been considerably less capital flow from the capital rich industrialised economies to the capital poor developing economies. Using a standard neo-classical growth model, Robert E. Lucas Jr. (1990) shows that the implied marginal productivity of capital in India is an astounding 58 times that of US. It is therefore a puzzle among economists why traditional theories cannot explain the capital flow (or lack thereof) from the developed to the developing economies.

There has been much research dedicated to explaining the 'Lucas Paradox'. Some economists have used differences in fundamentals (production structure, technology, policies, institutions) as explanations for the paradox. For example, Lucas cites the differences in human capital as the key reason why capital does not move to the South. On the other hand, other economists have mainly relied on capital market failures (expropriation risks, sovereign risks, asymmetric information) to resolve the paradox[21].

Even more interestingly, not all economists agree that there is a paradox in the first place. Carmen M. Reinhart and Kenneth S. Rogoff (2004) even suggest that the real paradox is that there is in fact too much capital flow to developing countries, considering the history and incidence of default in these economies. Aaron Tornell and Andres Velasco (1992) highlight the theoretical

[21] See Alfaro et al (2005) for a brief discussion on the various competing hypotheses.

possibility that poor property rights may even result in capital flight from a capital poor country to a capital rich country with better protection (or private access). It is therefore not abnormal that capital stays in the North as it offers better property rights. On a separate note, Abhijit Banerjee and Esther Duflo (2004) document comprehensive evidence which suggests that differences between the rates of return within some economies are larger than those across countries, which of course brings us to the question of whether it is relevant to focus on the Lucas Paradox.

Nevertheless, it is also clear from empirical research that it is often difficult to distinguish one theory from another. Countries with weak institutions tend to have lower human capital, and weak institutions tend to be associated with greater information asymmetry and expropriation risks. There can be too much or too little capital to the South, depending on which benchmark model is used, what instruments are used, what is defined as capital, and what kind of growth accounting is used[22].

Notwithstanding the various arguments presented, development over the past decade has necessitated a new understanding of the Lucas Paradox. The opening of China, India and other major emerging economies has resulted in increased flow of Foreign Direct Investment (FDI) to what is loosely termed as the South. The flow of capital is however highly uneven. In a recent working paper by Stephany Griffith-Jones and Jonathan Leape (2002), the authors highlight the huge differences in the capital flows to emerging economies. China attracted a fifth of all private capital flows to developing countries in

[22] Francesco Caselli and James Feyrer (2007) offer a similar insight by making a distinction between reproducible and non-reproducible capital. The authors argue that the reward to reproducible capital is in fact rather low in the South once proper accounting is done. There is therefore no paradox that capital does not move there.

the 1990s, peaking at $60 billion in 1997. India's share has been paltry by contrast, with a peak of $7 billion only in 1994. The latest figures show that China took in $72 billion in FDI in 2005, while India only received $6.6 billion[23]. Despite recent headline-grabbing growth rates from India, the FDI gap with China has not closed, although this might change in the near future.

If the Lucas paradox exists for India, it is on the face of it much less of a paradox for China. Is it therefore correct to conclude that China somehow has better fundamentals - institutions, technology, human capital, and/or less capital market imperfections? Given the fact that India is a stable parliamentary democracy, has a deeply entrenched English legal system with the associated emphasis on property rights, and a largely free press, it is difficult to turn the argument around and conclude that China has better institutions or better functioning markets that result in the huge difference in observed investment flows[24].

The puzzle is therefore not only why relatively little capital has flowed to the developing economies but also the distribution of the flow of capital to these economies. The objective of this paper is to synthesize the New

[23] China's cumulative inward FDI stands at $318 billion compared to $45 billion for India (UNCTAD). The difference in the levels of FDI is not due to differences in domestic investment. Inward FDI made up 11.3 per cent of China's gross capital formation between 1990 and 2000, but only 1.9 per cent compared to India. One of the explanations for the big difference is the effect of 'round-tripping' - domestic investment by Chinese firms disguised as FDI in order to gain a tax advantage. A look at the foreign investment position from the US however recorded the following difference: US cumulative investments in China and India (historical price) stand at $16.9B and $8.5B respectively. For manufacturing, the respective figures are $8.8B and $2.4B (Bureau of Economic Analysis). While 'round-tripping' may well account for some of the difference between the FDI that China and India have received, it is clear that China continues to receive signficantly more bona fide FDI than India.

[24] The problem with looking at historical data for defaults to explain current allocation, or predict future capital flows, becomes evident here. Historical data do not account for regime changes, changes in investor confidence and perception about the future. Reinhart and Rogoff (2004) duly note that India has never defaulted while China has defaulted on two occasions between 1901 and 2002. Yet it is still the case that China has taken the lion's share of FDI.

Economic Geography (NEG) understanding of the location of industries with more recent firm-heterogeneity trade models, in order to bring about a new understanding to an old puzzle as well as answer some of these new questions posed.

NEG researchers have had more than a decade of success in demonstrating how industrial agglomeration can result. These models demonstrate how a symmetric fall in trade costs can result in highly asymmetric outcomes (catastrophic agglomeration). The first NEG model, popularly known as the Core-Periphery (CP) model, by Paul R. Krugman (1991) demonstrates how the migration of industrial workers can result in industry concentration in a location. Subsequent work by Anthony J. Venables (1996) shows how vertical linkages (VL) between industries can result in firm migration with a similar agglomeration effect. These two models exhibit 'cumulative causation'[25]. An example of the mechanics is that firms locate where there are workers, and workers locate where there are firms (to reduce cost of living), giving rise to a feedback effect. These models tend to be highly intractable as a result. A more tractable model of industrial location is the 'Footloose Capital' (FC) model due to Philippe Martin and Carol Ann Rogers (1995). The key assumption of the model is that only capital is mobile, while workers and owners of capital are not. Capital income is costlessly repatriated, consumed locally. Since expenditure shares between the locations remain static regardless of the choice of industrial location, there is no agglomerative (or feedback) effect in this

[25] Some NEG models exhibit agglomeration - namely a feedback effect that generates ever increasing pressure for firms to locate in any one location. Other NEG models exhibit a concentration effect, where one observes more firms in one location even though there is no feedback effect. Both classes of NEG models explain the spatial locations of firms without appealing to technological externalities, which are captured by 'spillover' models.

class of models.

This paper has chosen to adopt the FC assumption. Firstly, international economics continue to be dominated by high capital mobility. Though there is international migration of labour, the speed at which adjustments take place is far slower, and its magnitude much smaller, compared to the movement of international capital. CP models with labour mobility might therefore be more useful in explaining regional adjustments within national economies than across countries. Secondly, the VL models rest on the assumption that firms require differentiated inputs or an Ethier type production function - specifically, all downstream firms using all upstream firms' inputs. While this offers a theoretical benchmark, its stylised assumption that all firms use all inputs is not often observed in reality.

In essence, the model in this paper assumes mobile capital, immobile labour, and firms with heterogeneous productivity. There are two locations, North and South. Differences between the two regions are characterised not by the aggregate production functions, but by differences in the productivity (pareto) distributions of firms. The shares of manufacturing firms in each location are then solved for in the equilibrium by equalising the ex-ante value of entry in both locations. Several new results emerge from the exercise.

Explaining the Lack of Capital Flow to South Firstly, while neoclassical models suggest that the productivity differences between North and South have to be very large to explain the lack of capital flow, this paper shows that a small improvement in North's productivity (by changing the mean of the pareto distribution) can have a dramatic impact on the share of firms, while keeping the returns to factors equal in both locations. This therefore provides an alternative resolution to the Lucas paradox. Admittedly, this paper does not explain why the small difference in productivity would arise in the first place. This question is better left to development or political-economy researchers [see James R. Tybout (2000) for a brief discussion].

Resolving The Paradox of Risk The second key result concerns the effect of risk. James R. Tybout (2000) for example notes that it is common to see very large plants existing side by side with very small ones in developing countries, even though there is little evidence to suggest plants in developing countries are inherently less productive. The author therefore suggests that this may be a result of 'uncertainty about policies . . . poor rule of law'. The assumption here is that the South has a riskier productivity draw.

A well known property of the profit function is its convexity. Consider the example of the constant elasticity of substitution (CES) preference function. For whatever the cost of production (inverse of productivity), the firm's revenue is bounded from below by zero - that is, revenue is always positive no matter how high the cost (and price) is. However, there is no upper bound to revenue. Therefore, a mean-preserving spread of productivity actually increases expected profits because of the very convexity of the profit function[26].

If the South were to have greater aggregate productivity risks while keeping its mean productivity equal to the North, this would imply that expected profit is higher there, and mobile capital will flow to the South until the expected return to capital is once again equalised for both locations. This is

[26] A mean-preserving spread of expenditure will have no such effect since it will still result in the same expected profits since the expenditure is homogeneous of degree one in expenditure.

the 'paradox of risk' for it contradicts commonplace intuition that firms shun locations perceived to have high risks to production. But in principle, the firm is a risk neutral entity. As long as the firm maximises expected profits, why does it care about risk?

It turns out that there is a good reason for this if one thinks of risk as outlined in a firm-specific productive risk in Marc J. Melitz (2003). Each firm will have to pay a sunk cost to attempt entry into a market. Upon the payment of this cost, the firm draws a level of productivity specific to itself, from an ex-ante distribution. The firm then makes the decision whether to continue production based on the level of realised productivity. If productivity is high enough, the firm will pay the fixed production cost and produce. Otherwise, the firm 'lets bygones be bygones' and exits.

It turns out that in equilibrium, the level of the sunk cost will have an impact on the location of industries. Suppose one location is riskier than the other while holding the mean of the productivity distribution constant. The riskier distribution will have fatter tails. Ceteris paribus, high sunk cost industries prefer to invest in less risky locations because the higher likelihood of entry dominates (the probability of a really bad draw is low). On the other hand, low sunk cost industries invest in higher risk locations (with fatter right side tails for productivity draws) since the chance of getting a really good productivity draw dominates. Given a particular sunk cost, a firm therefore has to balance these two effects. The model can explain why 'hi-tech' industries - characterised by high sunk costs - cluster in the less risky North while 'low-tech' industries move to the risky South.

The Importance of Trade Costs Finally in the standard FC model, if expenditures are equal in both locations, the distribution of firms will continue to be symmetric at all levels of iceberg trade costs except zero, at which the point of location of production is undefined. FC models can only achieve asymmetric concentration of industries through the home market effect (that is, different expenditure shares) whereby the location with the larger expenditure share has a more than proportionate share of firms. Without differences in expenditure shares, changes in trade costs cannot change industrial concentration.

The introduction of firm heterogeneity restores the potency of trade costs. Different levels of trade costs will result in different concentrations of industries, even if expenditures are the same in both locations.

3.2 The Model Setup

3.2.1 Endowments and Regions

There are two primary factors of production - capital and labour. There are two regions - North and South. The North has K_N units of capital and L_N units of labour while the South has K_S and L_S, all factors in fixed and known quantities. Capital is completely mobile between regions, and capital returns can be costlessly remitted to owners for consumption. Workers (who are also owners of capital) are completely immobile between regions, and their labour is supplied inelastically to the local market.

3.2.2 Preferences

There are two types of goods - agriculture (a) and manufacturing (m). The motivation is similar to most NEG models, with the agriculture sector equalising wages across economies in an equilibrium characterised by incomplete specialisation and without trade cost in agriculture. The j consumer's utility is given as

$$u_j = c_{mj}^{\mu} c_{aj}^{1-\mu}$$

where $c_{mj} = \left[\int_{\Omega} c_i^{\frac{\sigma-1}{\sigma}} di \right]^{\frac{\sigma}{\sigma-1}}$ is the consumption of the Ω set of manufactured goods, $\sigma > 1 > \mu > 0$.

3.2.3 Technology and Firms

Agriculture The agricultural sector has a constant returns to scale production function. For simplicity, units are chosen such that 1 unit of labour produces 1 unit of output. As per the usual assumption for NEG models, the agricultural good is costlessly traded between countries. This assumption equalises the price of the agricultural good between North and South, and also equalises the wage per unit of effective labour because of the perfectly competitive, constant returns to scale production.

Manufacturing The manufacturing sector requires a composite factor production κ which is produced by the primary factors - capital and labour - with a constant returns to scale Cobb-Douglas production technology

$$\kappa = AK^{\alpha}L^{1-\alpha}$$

where A is the aggregate technology parameter.

There is a large number of firms, each producing one variety. The firm's technology is homothetic and represented by the familiar increasing returns function

$$C_i = \left[f + \frac{q_i}{\varphi_i}\right] P_\kappa$$

where f is the fixed production cost and q the output. Therefore $\left[f + \frac{q_i}{\varphi_i}\right]$ gives the total input required of the firm in terms of κ and C is the total cost function given P_κ which is the price of the industrial composite. All firms have the same fixed cost but different levels of productivity φ.

Traditionally, the FC model has a disembodied technology - capital inputs for fixed cost and labour inputs for variable cost. Using a standard FC model but incorporating firm heterogeneity, Richard E. Baldwin and Toshihiro Okubo (2005) show how the home market effect can induce more productive firms to relocate to the larger market. That paper takes the ex-post productivity distribution of firms as given and ignores the entry or exit decision of firms. In a subsequent paper, Baldwin and Okubo (2006) introduce the entry and exit process. In that paper, the authors again highlight the home market effect, but further show how instantaneous entry and exit is a perfect substitute for relocation.

To achieve more analytical tractability, Baldwin and Okubo (2006) make some simplifying assumptions. Sunk cost, fixed export cost (beachhead cost), and variable production cost is borne by labour inputs only. Fixed production cost consists of capital only. The production technology is therefore a non-homothetic one, much like the standard FC model. In a firm heterogeneity setup however, there are many types of cost. Though it is not a criticism of the Baldwin and Okubo setup, it is not clear (at least theoretically) why the production technology should be so. Realistically, one could also think of sunk or beachhead cost to consist of capital only, or a combination of capital and labour.

This paper therefore adopts a more uniform approach towards the various types of costs by assuming a homothetic production technology that is more similar to Andrew B. Bernard, Stephen J. Redding and Peter K. Schott (2007) - known henceforth as BRS - where all costs require the same composition of inputs. There are several advantages with this setup.

Firstly, it is more realistic in that all costs will require capital and labour. The homotheticity of inputs towards manufacturing allows the model to be solved easily as in Melitz (2003) even in the presence of firm heterogeneity by making use of the 'Zero Cutoff Profits' and 'Free Entry' conditions. With a non-homothetic technology, this cannot be done. Secondly, changes in absolute endowments do not have an impact on firm level aggregates. Changes in endowments only affect the levels of composite as well as the capital-labour ratio. In a homothetic production setting, changes in endowments affect only the number of firms, relative returns of primary factors, and associated welfare, with no additional effect on firm level aggregates. The effect of changing endowments proportionately is just like changing market size. Consider the opposite case with a non-homothetic technology, supposing only capital is used for the sunk cost f_e. An increase in capital endowment, relative to labour, will mean that there will be relatively more resources for sunk cost compared to production. In equilibrium, it has to be more difficult to gain entry, and cutoff productivity has to increase. In other words, with a non-homothetic technology, changes in relative endowment will affect firm level aggregates.

Finally, though this paper draws inspiration from BRS (2007), there is a key difference. In BRS (2007) both factors of production - skilled and unskilled labour - are immobile. In this paper however, one of the factors - capital - is completely mobile. In essence, the technology function in this paper is a hybrid, combining elements of various research [Martins and Roger; Melitz; BRS] to incorporate various useful properties.

Capital Market This paper abstracts from any capital market imperfections by assuming that there is a well functioning capital market such that capital is transferred from owners to firms, and rewards are transferred costlessly back to owners for consumption.

Normalisation of Prices As the agriculture good is costlessly traded across the two regions, the wage rate w (yet to be solved) is therefore equalised between the two locations, assuming the consumption share of the manufactured good is small enough such that the agriculture sector continues to operate in both locations. Moreover, as capital is freely mobile across, the rental rate r is also equalised. The cost of the composite input κ - which depends on r and w - will therefore be also equalised between the two regions. Applying cost minimisation, together setting the P_κ as the numeraire, gives the following equation

$$P_\kappa = \frac{w^{1-\alpha} r^\alpha}{A} \left[\left(\frac{\alpha}{1-\alpha} \right)^{1-\alpha} + \left(\frac{1-\alpha}{\alpha} \right)^\alpha \right] \equiv 1 \tag{28}$$

This equilibrium relationship, in the situation of incomplete specialisation, allows the interest rate to be expressed in terms of wage rate and parameters (or vice versa). The advantage of choosing P_κ as the numeraire (rather than wages) is that it allows all equilibrium conditions for the manufacturing firms to be written in terms of κ only, without having to deal with the cost minimising price function of κ.

Furthermore, an implication of both cost minimisation and the equalisation of factor prices is that the rental-wage ratio can be expressed as

$$\frac{r}{w} = \left(\frac{\alpha}{1-\alpha}\right) \frac{L_M}{K_N + K_S} \tag{29}$$

where L_M is the total labour used in manufacturing[27]. Equation (29) allows r to be expressed as a function of w and parameters. Substituting this into equation (28), one can express the labour to capital ratio in manufacturing as a function of w only.

3.2.4 Pareto Productivity Distributions

All manufacturing firms face an ex-ante distribution of productivity in each location. This paper assumes pareto distributions for productivities in both North and South [Elhanan Helpman, Melitz and Stephen R. Yeaple (2005); BRS (2007); Baldwin and Okubo (2006)][28]. The parameters for the North are

[27] In an interior equilibrium, since r and w are common to both economies, they will have the same labour-capital ratios in the differentiated sectors. Hence $\frac{r}{w}K_{MN} = \frac{\alpha}{1-\alpha}L_{MN}$ and $\frac{r}{w}K_{MS} = \frac{\alpha}{1-\alpha}L_{MS}$ where K_{MN} and K_{MS} are the mobile capital deployed to the North and South respectively (while L_{MN} and L_{MS} are the labour employed in manufacturing respectively). Since all capital sums up to world endowment, $\frac{r}{w}(K_N + K_S) = \frac{\alpha}{1-\alpha}L_M$, where $L_M = L_{MN} + L_{MS}$.

[28] The relevant cumulative density, probability density, mean and variance are given as

$$G(\varphi) = 1 - \left(\frac{\varphi_m}{\varphi}\right)^k \qquad g(\varphi) = \frac{k\varphi_m^k}{\varphi^{k+1}}$$

$\bar{\varphi}_N$ and k_N, where $\bar{\varphi}_N$ specifies the minimum support and k_N the shape of the distribution. The corresponding parameters for the South are $\bar{\varphi}_S$ and k_S.

3.2.5 Sunk Cost

Firms trying to enter the manufactured goods market are required to pay a sunk cost of f_e (again in terms of κ) to draw the firm specific productivity φ. As capital is completely mobile, a firm can choose to pay this cost either in the North or in the South, upon which its productivity will be drawn from the respective distribution. The paper assumes that firms are not allowed to relocate their investment once they have selected on the initial location. The reason for this assumption is simple. Firm specific productivity is assumed to be tied to the institutional context in which sunk cost is incurred[29].

3.2.6 Trade Cost

Trade in the manufacturing sector is costly. There is a $\tau > 1$ iceberg trade cost for every unit shipped. In addition, exporters will have to incur a beachhead, or a fixed export cost f_X in order to export. Both costs are in terms of κ, paid in the home country. Selection into the export market will occur if there exist firms with productivity below φ that find it profitable to operate domestically (with domestic revenue r_D) but not export (thereby foregoing revenue r_X).

$$E(\varphi) = \frac{k\varphi_m}{k-1} \qquad Var(\varphi) = \frac{\varphi_m^2 k}{(k-1)^2(k-2)}$$

where $k > 2$ and $\varphi_m > 0$. For a pareto distribution, both mean and variance is decreasing in k.

[29]If both locations have the same ex-ante productivity distribution, no firms will relocate in equilibrium since the cutoffs are the same. An atomistic firm will have the same expected profits in both locations. If the productivity distributions are different, considering the effects of relocation requires an assumption to be made about whether productivity can be transferred across locations.

3.3 Trade Equilibrium Conditions

As usual, the agriculture sector equalises wages between the two locations

$$w = p_a = p_a^* = w^*$$

where Southern variables are denoted with the asterisk (except for variables related to productivity φ where locations are denoted with the subscript).

3.3.1 Export Partitioning

With CES preferences, the optimal pricing of a firm with productivity φ_1 is $p(\varphi_1) = \frac{\sigma}{\sigma-1}\frac{1}{\varphi_1}$, and the revenue given as $r(\varphi_1) = \frac{p(\varphi_1)^{1-\sigma}}{P^{1-\sigma}}E$, where E is the aggregate expenditure and P is the CES price aggregate. The ratio of revenues between two firms with productivities φ_1 and φ_2 can therefore be expressed as $\frac{r(\varphi_1)}{r(\varphi_2)} = \left(\frac{\varphi_1}{\varphi_2}\right)^{\sigma-1}$. Furthermore, one can define a firm with cutoff productivity φ^* as the marginal firm - one that just makes enough operating profits to cover the fixed cost of production f. This firm therefore satisfies the relationship of net operating profits equalling the fixed cost: $\frac{1}{\sigma}r(\varphi^*) = f$. This allows one to write the revenue of a firm with an average productivity of $\tilde{\varphi}$ (to be defined later) as a function of the cutoff productivity φ^* only

$$r(\tilde{\varphi}) = \left(\frac{\tilde{\varphi}}{\varphi^*}\right)^{\sigma-1} f$$

Average profits from domestic sales become

$$\bar{\pi}_D = \left(\frac{\tilde{\varphi}}{\varphi^*}\right)^{\sigma-1} f - f = \left[\left(\frac{\tilde{\varphi}}{\varphi^*}\right)^{\sigma-1} - 1\right] f$$

Analogously, profits from exporting become

$$\bar{\pi}_X = \left[\left(\frac{\tilde{\varphi}_X}{\varphi_X^*}\right)^{\sigma-1} - 1\right] f_X$$

where φ_X^* is the export cutoff (greater than φ^*) because of the exporting partition condition which is assumed to hold (that is, not all firms export), and $\tilde{\varphi}_X$ is the average productivity of exporters.

3.3.2 Average Profits

Given these standard derivations, the average profits in the North can be written as

$$\bar{\pi} = \left[\left(\frac{\tilde{\varphi}_N}{\varphi_N^*}\right)^{\sigma-1} - 1\right] f + \tilde{p}_X \left[\left(\frac{\tilde{\varphi}_{NX}}{\varphi_{NX}^*}\right)^{\sigma-1} - 1\right] f_X \tag{30}$$

where φ_N^* is the cutoff productivity for entry, $\tilde{\varphi}_N$ the average productivity of all Northern firms above the cutoff, φ_{NX}^* the cutoff productivity into export, and $\tilde{\varphi}_{NX}$ is the average productivity of Northern exporters. Since only those manufacturers with a productivity draw greater than φ_{NX}^* can export, the term $\tilde{p}_X = \left(\frac{\varphi_N^*}{\varphi_{NX}^*}\right)^{k_N}$ gives the conditional probability of having a high enough productivity to export, conditional upon entry. In short, conditioned on successful entry, the first term on the right hand side gives the expected profits from domestic sales while the second term gives the expected profits from exporting.

The analogous expression for the South is

$$\bar{\pi}^* = \left[\left(\frac{\tilde{\varphi}_S}{\varphi_S^*}\right)^{\sigma-1} - 1\right] f + \tilde{p}_X^* \left[\left(\frac{\tilde{\varphi}_{SX}}{\varphi_{SX}^*}\right)^{\sigma-1} - 1\right] f_X \tag{31}$$

where $\tilde{p}_X^* = \left(\frac{\varphi_S^*}{\varphi_{SX}^*}\right)^{k_S}$ is the conditional probability of exporting in the South.

The marginal firms in the North and South, with productivities φ_N^* and φ_S^*, recover only the fixed cost of production f in equilibrium. This gives the following relationship

$$\left(\frac{\sigma}{\sigma-1}\frac{1}{\varphi_N^*}\right)^{1-\sigma}\left[\frac{\mu E}{P^{1-\sigma}}\right] = \sigma f = \left(\frac{\sigma}{\sigma-1}\frac{1}{\varphi_S^*}\right)^{1-\sigma}\left[\frac{\mu E^*}{P^{*1-\sigma}}\right] \qquad (32)$$

These are effectively zero profit conditions that will help pin down the productivity cutoffs in equilibrium.

3.3.3 Productivities of Northern and Southern Firms

As with the usual derivations in such models, average productivities of Northern and Southern firms - $\tilde{\varphi}_N$ and $\tilde{\varphi}_S$ - are functions of the respective cutoffs only[30]. The pareto productivity distributions allow the ratios between the average productivities and their respective cutoffs to be written as a function of parameters only $\left(\frac{\tilde{\varphi}_N}{\varphi_N^*}\right)^{\sigma-1} = \left(\frac{\tilde{\varphi}_{NX}}{\varphi_{NX}^*}\right)^{\sigma-1} = \left[\frac{k_N}{k_N+1-\sigma}\right]$, with analogous expressions holding for the South.

Together, these properties give the extremely useful result that $\frac{\tilde{\varphi}_N}{\varphi_N^*} = \frac{\tilde{\varphi}_{NX}}{\varphi_{NX}^*}$ with the pareto productivity distributions. Though exporters have a higher average productivity, the ratio of average productivity of all producers to the entry cutoff is exactly the same as the ratio of average productivity of all exporters to the export cutoff. Plugging these conditions into equations (30) and (31) greatly simplifies these expressions and the characterisation of

[30] These are $\tilde{\varphi}_N = \left[\frac{1}{1-G_N(\varphi_N^*)}\int_{\varphi_N^*}^{\infty}\varphi^{\sigma-1}g_N(\varphi)d\varphi\right]^{\frac{1}{\sigma-1}}$ and $\tilde{\varphi}_S = \left[\frac{1}{1-G_S(\varphi_S^*)}\int_{\varphi_S^*}^{\infty}\varphi^{\sigma-1}g_S(\varphi)d\varphi\right]^{\frac{1}{\sigma-1}}$. With pareto productivity distributions, these can be further simplified to $\tilde{\varphi}_N = \left[\frac{k_N}{k_N+1-\sigma}\right]^{\frac{1}{\sigma-1}}\varphi_N^*$ and $\tilde{\varphi}_S = \left[\frac{k_S}{k_S+1-\sigma}\right]^{\frac{1}{\sigma-1}}\varphi_S^*$.

equilibrium.

Finally, a firm with φ_N^* makes zero profits in the domestic market, while a firm with φ_{NX}^* makes zero profits from exporting (with the analogous relationships holding for the South as well)[31].

3.3.4 Aggregate Productivity and Prices

The aggregate productivity and price level in a location depend not only on domestic firms, but also on foreign firms selling there. Define the total number of varieties in the North by $M = n + \tilde{p}_X^* n^*$. This indicates that the number of varieties in the North is made up of n domestic firms and $\tilde{p}_X^* n^*$ of Southern firms that are successful in exporting to the North. The corresponding expression for the South is $M^* = \tilde{p}_X n + n^*$.

The average productivity of the North becomes the weighted average of productivities of Northern firms and Southern exporters

$$\hat{\varphi} = \left[\frac{1}{M}\left(n\tilde{\varphi}_N^{\sigma-1} + \tilde{p}_X^* n^* \phi \tilde{\varphi}_{SX}^{\sigma-1}\right)\right]^{\frac{1}{\sigma-1}} \tag{33}$$

where $\phi = \tau^{1-\sigma}$ is the freedom of trade index. The corresponding equation for the South can be written as

$$\hat{\varphi}^* = \left[\frac{1}{M^*}\left(\tilde{p}_X n \phi \tilde{\varphi}_{NX}^{\sigma-1} + n^* \tilde{\varphi}_S^{\sigma-1}\right)\right]^{\frac{1}{\sigma-1}} \tag{34}$$

With these definitions of productivities, the aggregate price levels in the North

[31] When the countries are symmetric, the respective export cutoffs are a function of production cutoffs and parameters only, with $\varphi_{NX}^* = \varphi_N^* \tau \left(\frac{f_X}{f}\right)^{\frac{1}{\sigma-1}}$ and $\varphi_{SX}^* = \varphi_S^* \tau \left(\frac{f_X}{f}\right)^{\frac{1}{\sigma-1}}$. When the countries are not symmetric, one can show that $\varphi_{NX}^* = \tau \left(\frac{f_X}{f}\right)^{\frac{1}{\sigma-1}} \varphi_S^*$ and $\varphi_{SX}^* = \tau \left(\frac{f_X}{f}\right)^{\frac{1}{\sigma-1}} \varphi_N^*$ [see Svetlana Demidova (2006)]

and South are given as

$$P = M^{\frac{1}{1-\sigma}} \frac{\sigma}{\sigma - 1} \frac{1}{\hat{\varphi}} \qquad P^* = M^{*\frac{1}{1-\sigma}} \frac{\sigma}{\sigma - 1} \frac{1}{\hat{\varphi}^*} \tag{35}$$

This completes the characterisation of the aggregate price levels for both locations. The aggregate prices P and P^* in equation (35) can also be substituted into the marginal firm conditions in equation (32), allowing the zero profit conditions to be expressed as function of firm mass and productivity cutoffs only.

3.3.5 Equalisation of Expected Values of Entry in North and South

Free entry ensures that the ex-ante value of entry must be equal for both locations if there is to be an interior solution (with manufacturing firms in both locations) The condition for an interior equilibrium can be written as $\tilde{p}\tilde{\pi}_N = \tilde{p}^*\tilde{\pi}_S^* = f_e$, where $\tilde{p} = 1 - G_N(\varphi_N^*) = \left(\frac{\bar{\varphi}_N}{\varphi_N^*}\right)^{k_N}$ and $\tilde{p}^* = 1 - G_S(\varphi_S^*) = \left(\frac{\bar{\varphi}_S}{\varphi_S^*}\right)^{k_S}$ are the entry probabilities of the North and South respectively[32]. With the appropriate substitutions, this expression can be explicitly written as

$$\tilde{p}\left(\frac{\sigma - 1}{k_N + 1 - \sigma}\right)(f + \tilde{p}_X f_X) = f_e$$

$$\tilde{p}^*\left(\frac{\sigma - 1}{k_S + 1 - \sigma}\right)(f + \tilde{p}_X^* f_X) = f_e \tag{36}$$

where $\tilde{p}_X$ and $\tilde{p}_X^*$ are the conditional probabilities of exporting.

[32] If manufacturing concentrates completely in one location, one of these equalities will not hold. Expected profits in one location do not cover the sunk cost f_e in equilibrium and no manufacturing firms locate there. This can be used to pin down the break/sustain point.

3.3.6 Market Clearing

There are in equilibrium n successful entrants in the North and n^* in the South. But due to the cutoffs, the number of firms that attempt entry has to be higher. The total number of firms that attempt entry, including those below the cutoffs, is

$$n_e = \frac{n}{\tilde{p}} \qquad\qquad n_e^* = \frac{n^*}{\tilde{p}^*}$$

where n_e and n_e^* are the total number of entry attempts in the North and South respectively.

The composite input κ is used for four purposes - sunk cost (f_e), fixed production cost (f), marginal production cost, and export costs (this is incurred by exporters only). The key to note here is that even unsuccessful entrants will use up industrial inputs. The marginal cost for each firm is $\frac{1}{\varphi}$, a firm-specific variable. The aggregate variable production cost in the North can be written as $n\left(\frac{k_N}{k_N+1-\sigma}\right)(\sigma-1)f$ [see Appendix B:3.8.2]. Aggregate composite input used in the North becomes

$$\kappa = n\left\{f + \left(\frac{k_N}{k_N+1-\sigma}\right)(\sigma-1)f + \frac{f_e}{\tilde{p}} + \tilde{p}_X\left[f_X + \left(\frac{k_N}{k_N+1-\sigma}\right)(\sigma-1)\frac{f_X}{\tau}\right]\right\} \tag{37}$$

Multiplied by the number of firms, the first term within the brackets on the right hand side is the total fixed production cost. The second term on the right (again multiplied by the number of firms n) is the aggregate variable cost of all firms. The third term (multiplied by the number of firms) is the total sunk cost incurred, including that of the unsuccessful firms. Finally, the terms inside the square brackets (multiplied by the number of firms) are the total beachhead and exporting production costs, which are incurred by exporters only. An analogous term can be written for the South

$$\kappa^* = n^* \left\{ f + \left(\frac{k_S}{k_S + 1 - \sigma} \right) (\sigma - 1) f + \frac{f_e}{\tilde{p}^*} + \tilde{p}_X^* \left[f_X + \left(\frac{k_S}{k_S + 1 - \sigma} \right) (\sigma - 1) \frac{f_X}{\tau} \right] \right\} \tag{38}$$

The above two expressions therefore give the quantity of the composite input κ demanded in the North and South respectively.

Due to the cost minimisation property, the derived demand for capital is $K = \frac{1}{A} \left(\frac{w}{r}\right)^{1-\alpha} \left(\frac{\alpha}{1-\alpha}\right)^{1-\alpha} \kappa$. By substituting the demands of κ into the appropriate conditional demands, one can derive the demands of the primary factors capital and labour. Since the total demand of capital in the world must be equal to the endowment, the capital clearing condition can be written as

$$\bar{K}_W = \frac{1}{A} \left(\frac{w}{r} \right)^{1-\alpha} \left(\frac{\alpha}{1-\alpha} \right)^{1-\alpha} (\kappa + \kappa^*) \tag{39}$$

Equation (39) converts the industrial inputs into capital by substituting κ in equations (37) and (38) into the appropriate cost-minimising function. This is the first market clearing equation.

Similarly, since the conditional demand for labour (for manufacturing) can be written as $L = \frac{1}{A} \left(\frac{r}{w}\right)^{\alpha} \left(\frac{1-\alpha}{\alpha}\right)^{\alpha} \kappa$, the total labour requirement for manufacturing becomes

$$L_M = \frac{1}{A} \left(\frac{r}{w} \right)^{\alpha} \left(\frac{1-\alpha}{\alpha} \right)^{\alpha} (\kappa + \kappa^*) \tag{40}$$

As labour is also used for agriculture, the total manufacturing labour does not equal to the total labour endowment. Instead, the amount of labour available for agriculture is whatever labour not used in manufacturing. This has to be equal to the real demand for agricultural goods (nominal expenditure divided by the price of agriculture goods, which is w), giving the agricultural market (or labour market) clearing condition

$$\bar{L}_W - L_M = (1-\mu)\left[\frac{E+E^*}{w}\right] \tag{41}$$

Substituting equation (40) into (41) then provides the second market clearing condition. With CES preferences, the manufacturing goods market clears since the expenditure on each firm is equal to its revenue. With Walras's Law, the agriculture market also clears.

3.3.7 Aggregate Expenditure

As owners of capital are immobile, all capital returns are remitted to the owners and consumed locally. The aggregate expenditures for the North and South are simply their respective factor endowments multiplied by the rental and wage rates, which are the same across countries in the incomplete specialisation equilibrium

$$E = rK_N + wL_N \qquad\qquad E^* = rK_S + wL_S$$

3.3.8 Equilibrium Solution

The endogenous variables for equilibrium are $\{w, \varphi_N^*, \varphi_S^*, n, n^*\}$ - although the interest rate is endogenous, it can be recovered by equation (28). For the five endogenous variables, the equilibrium is pinned down (after appropriate substitutions) by (i) two ex-ante free entry conditions in equation (36); (ii) zero profit condition in equation (32); and two market clearing conditions in equation (39) and (41).

3.3.9 Solving for Global Manufacturing Labour

From equation (41), the global production of agriculture is

$$\begin{aligned}
\bar{L}_W - L_M &= (1-\mu)\left[\frac{E+E^*}{w}\right] \\
&= (1-\mu)\left[\frac{r\bar{K}_W + w\bar{L}_W}{w}\right] \\
&= (1-\mu)\left[\frac{r}{w}\bar{K}_W + \bar{L}_W\right]
\end{aligned}$$

The second equality makes use of the fact that the global expenditure is a function of wage-rental and global endowments $E+E^* = r\bar{K}_W + w\bar{L}_W$. Substituting the rental-wage ratio from equation (29) then allows the world's labour employed in manufacturing to be expressed as a function of endowments and parameters only

$$L_M = \left[\frac{\mu(1-\alpha)}{1-\alpha\mu}\right]\bar{L}_W \tag{42}$$

Note that μ (which is the share of manufacturing in consumption) has to be less than 1 for $\left[\frac{\mu(1-\alpha)}{1-\alpha\mu}\right]$, the share of global labour in manufacturing, to also be less than 1.

3.3.10 Solving for Rental-Wage Ratio

Substituting equation (42) back to equation (29) then allows the rental-wage ratio to be expressed as a function of parameters only

$$\frac{r}{w} = \left[\frac{\alpha\mu}{1-\alpha\mu}\right]\frac{\bar{L}_W}{\bar{K}_W} \tag{43}$$

Note that the rental-wage ratio is also unaffected by any firm level variables. It depends on the endowments ratio and parameters only.

3.3.11 Solving for Total Composite Resource

Equation (40) gives the relationship between L_M and the composite resource $\kappa + \kappa^*$

$$L_M = \frac{1}{A}\left(\frac{r}{w}\right)^{\alpha}\left(\frac{1-\alpha}{\alpha}\right)^{\alpha}(\kappa + \kappa^*)$$

Substituting equations (42) and (43) into the above will give

$$\kappa + \kappa^* = A(1-\alpha)^{1-\alpha}\alpha^{\alpha}\left[\frac{\mu}{1-\alpha\mu}\right]^{1-\alpha}\bar{L}_W^{1-\alpha}\bar{K}_W^{\alpha} \tag{44}$$

The total composite resource available to the manufacturing sector is an increasing function of endowments, aggregate technology A, and share of manufacturing consumption μ (because this reduces the amount of labour required for agriculture). This therefore pins down the total composite factor supply in terms of endowments and parameters only.

3.4 When the North Is More Productive

From the free entry conditions in equation (36),

$$\tilde{p}\left(\frac{\sigma-1}{k_N+1-\sigma}\right)(f+\tilde{p}_X f_X)=\tilde{p}^*\left(\frac{\sigma-1}{k_S+1-\sigma}\right)(f+\tilde{p}_X^* f_X)$$

Since $k=k_N=k_S$, this condition becomes

$$\tilde{p}(f+\tilde{p}_X f_X)=\tilde{p}^*(f+\tilde{p}_X^* f_X)$$

Writing this equation more explicitly

$$\left(\frac{\bar{\varphi}_N}{\varphi_N^*}\right)^k\left[f+\left(\frac{\varphi_N^*}{\varphi_{NX}^*}\right)^k f_X\right]=\left(\frac{\bar{\varphi}_S}{\varphi_S^*}\right)^k\left[f+\left(\frac{\varphi_S^*}{\varphi_{SX}^*}\right)^k f_X\right]$$

From here, the paper states a few simplifying relationships. First, $\bar{\varphi}_N=\psi\bar{\varphi}_S$ where $\psi>1$ represents the rightward shift of North's support for the productivity distribution. Second, one can make use of the following relationships $\varphi_{NX}^*=\tau\left(\frac{f_X}{f}\right)^{\frac{1}{\sigma-1}}\varphi_S^*$ and $\varphi_{SX}^*=\tau\left(\frac{f_X}{f}\right)^{\frac{1}{\sigma-1}}\varphi_N^*$ when solving for cutoffs for asymmetric countries where $Z=\tau\left(\frac{f_X}{f}\right)^{\frac{1}{\sigma-1}}$ [see Demidova (2006)]. Third, let $\eta^k=\frac{f_X}{f}$ be the ratio of the two fixed cost. Using these simple relationships, the above equation can be further simplified to

$$\left(\frac{\psi}{\varphi_N^*}\right)^k\left[\frac{fZ^k\varphi_S^{*k}+\varphi_N^{*k}f_X}{Z^k\varphi_S^{*k}}\right]=\left(\frac{1}{\varphi_S^*}\right)^k\left[\frac{fZ^k\varphi_N^{*k}+\varphi_S^{*k}f_X}{Z^k\varphi_N^{*k}}\right]$$

After cancellations of terms, one can simplify this to

$$\psi^k\left(Z^k\varphi_S^{*k}+\varphi_N^{*k}\eta^k\right)=Z^k\varphi_N^{*k}+\varphi_S^{*k}\eta^k$$

By grouping the terms, one can express one cutoff as a function of another

$$\varphi_N^* = \left[\frac{\eta^k - \psi^k Z^k}{\psi^k \eta^k - Z^k}\right] \varphi_S^* = \nu \varphi_S^* \tag{45}$$

where $\nu = \left[\frac{\eta^k - \psi^k Z^k}{\psi^k \eta^k - Z^k}\right] > 1$ is simply a function of parameters only.

3.4.1 Solving for Cutoffs

This subsection solves for South's productivity cutoff using the free entry condition

$$f_e = \tilde{p}^* \left(\frac{\sigma - 1}{k_S + 1 - \sigma}\right)(f + \tilde{p}_X^* f_X)$$

Since the probability of entry is given as $\tilde{p}^* = \left(\frac{\bar{\varphi}_S}{\varphi_S^*}\right)^k$ and the conditional probability of export $\tilde{p}_X^* = \left(\frac{\varphi_S^*}{\varphi_{SX}^*}\right)^k = \left(\frac{\varphi_S^*}{Z\varphi_N^*}\right)^k$, the above equation becomes

$$f_e = \left(\frac{\bar{\varphi}_S}{\varphi_S^*}\right)^k \left(\frac{\sigma - 1}{k + 1 - \sigma}\right)\left[f + \left(\frac{1}{\nu}\right)^k f_X\right] \tag{46}$$

This gives the analytical closed-form solution to South's cutoff φ_S^*. The Northern cutoff can be derived from equation (45). With these productivity cutoffs, the export cutoffs can also be derived. The break point - where all firms locate in the North - is characterised in the Appendix.

3.4.2 Aggregate and Firm-Level Variables

In characterising this equilibrium, a few facts stand out. Firstly, the equilibrium rental-wage ratio in equation (43) is unaffected by any firm level variables. The amount of industrial resources $\kappa + \kappa^*$ in equation (44) available for the differentiated sector is also independent of firm level variables. These are all functions of endowments and other parameters only. As mentioned before, symmetric changes in endowments (relative or absolute), therefore do not have any impact on firm level variables. Secondly, firm level productivity cutoffs are solved through the free entry conditions in equation (36), and are also completely independent from interest or wage rates. The only interaction between firm-level and aggregate variables is how the size of resources $\kappa + \kappa^*$ affect the number of firms in equilibrium.

This is the key property of the homothetic production function. The same intuition is highlighted by Melitz (2003) with a single factor of production, where the size of the market affects only the number of firms in equilibrium, not firm level aggregates. Though North and South have different cutoffs due the difference in the productivity distributions, changes in endowments do not affect respective cutoff productivities or average profits. To the atomistic firm, the supply of factors is completely elastic. The size of the endowment will determine only the number of firms in equilibrium. Relative endowment will affect the $\frac{r}{w}$ ratio required to clear the respective markets, but otherwise will also have no impact on firm level aggregates.

3.4.3 A Numerical Example

This subsection provides a simple numerical example to illustrate the equilibrium characterised[33]. This paper does not make any empirical estimates on any parameters. Instead, parameters on preferences and pareto distribution are taken from existing research. The choice of endowment is arbitrary. How-

[33]Numerical solutions are obtained through MATLAB. An initial estimate is provided for all the variables. The endogenous variables are then solved through the equilibrium conditions, and incremental updates in each round are carried out by taking the weighted average between the 'old' and 'new' solutions, until there are no further changes (convergence). The solution method is similar to Krugman (1991) and BRS (2007). I am grateful to Stephen Redding for sharing the MATLAB codes.

ever, the same level of endowment is chosen for the North and South in order not to introduce the home market effect that would otherwise be evident in an Economic Geography model. This assumption will be relaxed later to bring out the home market effect. The list of parameters is provided in Appendix B:3.8.1. The set of cost parameters is $\{f, f_e, f_X, \tau\}$. These parameters will also be varied in various numerical solutions to highlight the effects of changes in them.

In the first set of numerical solutions, North and South have the same distribution shape $k_N = k_S = 3.6$. However, North is given a better productivity compared to the baseline scenario, $\bar{\varphi}_N = 0.205 > \bar{\varphi}_S = 0.2$. This shifts the North's distribution rightwards (first degree stochastic dominance). The North is 2.5 per cent more productive than the South on the basis of the unconditional mean.

Even though North and South have the same level of expenditure (given the same level of endowment), the slight perturbation of the pareto distributions results in dramatic differences in industry location. The equilibrium effects on industrial concentration are presented in Table 5 for three different levels of trade cost (the Tomahawk diagram will be presented in a later section).

Table 5: Share of Firms and Capital in More Productive North

$f = 10$ $f_e = 10$ $f_X = 10$	$\tau = 1.40$	$\tau = 1.30$	$\tau = 1.20$
Share of Firms	0.535	0.566	0.661
Share of Capital	0.550	0.590	0.700

The firm-level variables with a relatively low level of trade cost $\tau = 1.10$ are presented in Table 6.

Table 6: Equilibrium Variables with More Productive North

For $\tau = 1.20$	**North**	**South**
Cutoff Productivity	0.3358	0.3104
Probability of Successful Entry	0.1693	0.2054
Average Firms' Productivity, on entry	0.5745	0.5312
Aggregate Price Levels	0.1331	0.1440

The results show that a reasonably small perturbation in the productivity distribution in the North can have a significant impact on the location of firms and capital. A 2.5 per cent increase in the unconditional mean of the productivity distribution creates a high concentration of industrial activity in the North at a intermediate-low level of trade cost of $\tau = 1.20$ (Table 5). The intuition becomes clear in Table 6. The better productivity distribution in the North means that firms there are more productive and profitable. More firms need to move there until the effects of local market competition cancel out any productivity advantages.

Another striking feature of this equilibrium is that in an interior equilibrium r and w are in fact the same in both locations, despite a higher level of capital in the North. The South continues to have a lower aggregate $\frac{K}{L}$ ratio compared to the North, but the marginal returns to capital is the same in both North and South. The Lucas paradox disappears. The superiority of the North is not in the aggregate production function, but is due to an improvement in firm-specific productivity draws.

Thirdly, the fall in trade cost will accentuate the advantages of locating in the North even though the levels of expenditure are the same in each location. In the traditional FC model, if the expenditures of both locations are the same, location of firms will be symmetric at all positive levels of trade cost. The concentration of industry depends on the home market effect. In other words, trade cost is completely 'impotent' in creating asymmetric concentration when the two markets are of equal size.

This is however not the case here. Expenditure is the same in both locations, but the fall in trade cost brings about an increasing concentration of industry to the North. The key to understanding this lies in the inspection of equations (33) and (34). Because the North has a superior productivity distribution, its firms are more productive in equilibrium. In autarky, North and South's CES price indices only reflect the productivities of their domestic firms.

Therefore, with the opening to trade and the fall in trade cost, the increase in ϕ creates a greater increase in weighted average productivity in the South $\hat{\varphi}^*$ compared with $\hat{\varphi}$. Competitive pressure intensifies more quickly in the South with a fall in trade cost, thereby accentuating the advantages of locating in the North. Conversely, Northern firms are less affected by the effects of increased competition as a result of freer trade since they are more productive than their Southern counterparts.

3.5 The Impact of Risk

In the previous sub-section, the North is more attractive due to its better productivity distribution. However, suppose the South is not less productive but riskier. How will this change the distribution of capital and firms? It is important that the impact of risk is clearly understood since one of the competing hypotheses on why relatively little capital flows to the South is the inherent riskiness in investing there (expropriation risk, political risk etc). In this set of numerical solutions, it is precisely this effect that is being modelled by allowing the two productivity distributions to have the same mean productivity but greater dispersion in the South.

In this set of numerical solutions, the North has the following minimum support $\bar{\varphi}_N = 0.205 > \bar{\varphi}_S = 0.2$. Moreover, the shape of the North's distribution is tighter with $k_N = 3.8 > k_S = 3.6$. The result of this is that the unconditional productivity means in both locations are the same with $c_N = c_S = 0.277$. However, the variance in the North is 16 per cent smaller than the South. The set of parameters in fact creates a 'mean preserving spread' of the productivity distribution in the South. The South is not less productive on average, but has higher risk as characterised by the higher variance. The numerical solution to the equilibrium firm shares, with a moderate level of trade cost $\tau = 1.2$ and different levels of sunk cost f_e, are presented in Table 7.

Table 7: Share of Firms and Capital in Less Risky North with Different Sunk Cost

$\tau = 1.30$ $f_e = 10$ $f_X = 10$	$f_e = 5$	$f_e = 10$	$f_e = 30$
Share of Firms	0.288	0.335	0.396
Share of Capital	0.196	0.244	0.309

The firm-level variables with $\tau = 1.20$ and $f_e = 30$ are presented in Table 8.

Table 8: Equilibrium Variables with Riskier South

For $\tau = 1.30$ and $f_e = 30$	**North**	**South**
Cutoff Productivity	0.2121	0.2370
Probability of Successful Entry	0.8629	0.5422
Average Firms' Productivity, upon entry	0.3416	0.4057
Aggregate Price Levels	0.2108	0.1886

The results of this sub-section show the effects of greater variance in the productivity distribution. There is a tendency for industrial concentration in the South. The higher variance in the South implies that there is a fatter right side tail for the pareto distribution. As can be seen from Table 8, the effect of this is that although the probability of entry is lower in the South, the average productivity upon successful entry is in fact higher in the South due to the fatter right tail.

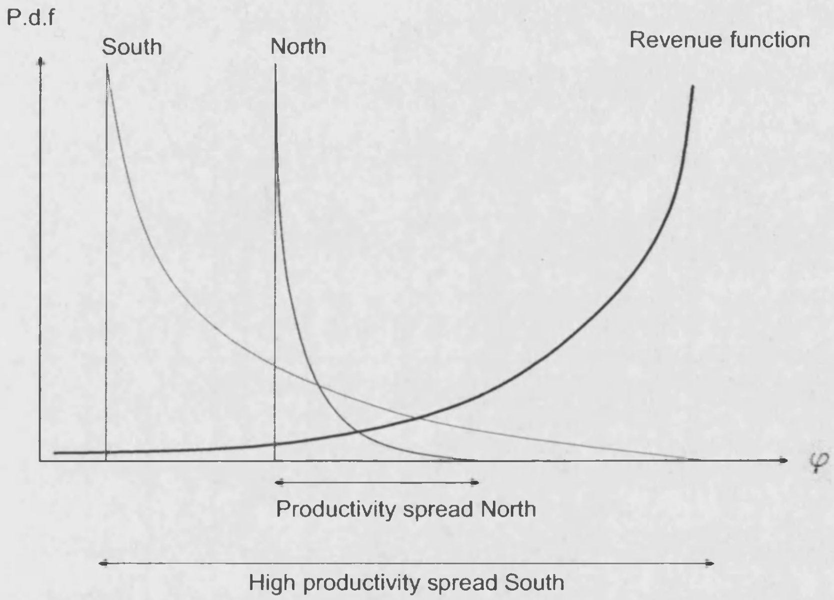
Proposition 6 *When expenditures are equal in both locations, increasing sunk cost f_e will result in greater share of industry for the North.*

Proof. See Appendix B:3.8.4. ∎

What is the economic intuition here? After a firm invests in the sunk cost and discovers its productivity, it can decide whether to incur the fixed production cost f. Incurring the sunk cost creates an option whether to produce, as a firm has a choice of whether to carry out production. At low values of the sunk cost, the South is more attractive since it offers a greater probability of a high productivity draw (and higher average productivity). At higher values of the sunk cost however, this option effectively becomes more expensive and reduces the attraction of the South.

To understand the impact that cutoffs have on distribution of capital, it is useful to first think of ex-ante entry conditions without cutoffs. Suppose a firm has to make a decision to enter either the North or South market in one stage. In other words, there is no separation of f_e and f - a firm discovers its productivity and can begin production without any further investment. Further suppose that the South has a higher productivity spread. With the CES demand, the revenue function is always bounded from below by zero but has no upper bound. A higher productivity spread in the South in fact increases the ex-ante profits, thereby drawing more firms there until any ex-ante difference is equalised. This is the effect seen in Figure 3, where the same CES revenue function is superimposed on the probability densities of the North and South's productivity distribution.

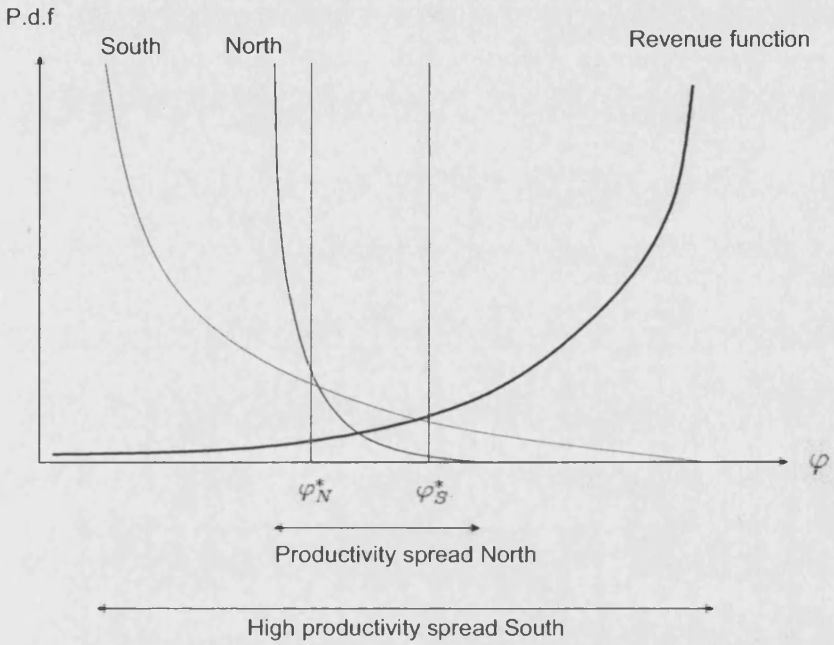
Figure 3: Effects on Expected Revenue with Different North-South Productivity Distributions



The narrow right side tail of the North means that it is giving up the potential for high productivity draws and high profits. Because of the convexity of the revenue function, a firm in the North will have lower expected profits, ceteris paribus. If North and South have similar expenditures, more firms will have to locate to the South until profits are equalised.

However, given a two stage entry game (f_e to discover productivity and f to produce), the riskier location can imply a smaller probability of entry. With the cutoff productivity in a two stage entry decision, revenue functions are now truncated left of the cutoff (see Figure 4). Given that the two locations have different productive distributions, the effect is asymmetric.

Figure 4: Effects of Cutoffs on Expected Revenue (Truncation)



The revenue function is truncated (falls to zero) left of the respective cutoffs. The probability of successful entry can become higher in the North, dominating any foregone probability of an high probability draw in the South. If that happens, more firms will have to locate to the North. It is also possible that potential for high productivity draws in the South to dominate the higher entry probability in the South, and more firms locate to the South in that case. Expected profits are determined by two components - firstly the probability of successful entry and secondly, the expected productivity and profitability post-entry. It is the balance of these two margins that changes the relative attractiveness of each location.

Consider then the effect of the sunk cost f_e. A higher f_e will always shift the cutoffs to the left while a lower f_e shifts cutoff rightwards. As f_e increases and cutoffs shift leftwards, the probability of successful entry always rises faster in the North since it has a narrower productivity distribution. Conversely as f_e falls, North's entry probability falls faster than the South's. The level of f_e therefore changes the balance of the two margins affecting a firm's decision on where to locate. Ignoring the effect of market size for the moment and keeping expenditures the same in both locations, increasing f_e will increase the expected profits of North and result in more firms locating there, and vice versa [see Appendix B:3.8.4].

3.6 Extension to Multi-Industry and Larger North

In this section, the paper further generalises the results to an economy with more than one differentiated industry. As before, there are two regions North and South - where both have the same mean productivity, but South is riskier. The productivity distributions are the same as the previous section[34]. What is different here is that there are two differentiated sectors, A and B. The consumption shares are identical at $\mu = 0.15$ (this is kept small so that the agriculture sector continues to operate in both locations). There are no inter-industry linkages. Furthermore, the North is given an endowment advantage - its capital and labour endowment are 20 per cent more than the South - roughly in line with the idea that developed markets are bigger in size. The paper then shows using numerical solutions how market size can interact with

[34] Where $\bar{\varphi}_N = 0.2077 > \bar{\varphi}_S = 0.2$ and $k_N = 4 > k_S = 3.6$.

the level of sunk costs to result in different types of specialisation as trade becomes freer.

The two differentiated sectors have exactly the same industrial structure except for one difference. Industry A is a high-tech industry with a sunk cost of $f_e(A) = 30$, while industry B is a low-tech industry with a sunk cost of $f_e(B) = 1$. The two industries have exactly the same cost structure otherwise with $f = 10$ and $f_X = 10$. They also have the same iceberg trade cost. These assumptions are not meant to be realistic. For example, industries with lower sunk cost (low-tech) tend to have higher elasticity of substitution. The assumptions are kept as simple as possible here, only for the purpose of illustrating how two industries with different f_e can end up concentrating at different locations with different ex-ante productivity distributions.

3.6.1 Tomahawk Diagram

The paper has thus far not presented any Tomahawk diagrams since all intuition will be captured in this section. In the diagram, the level of trade cost falls from the left to the right in the X-axis (ϕ increases from 0 to 1). The Y-axis are the shares of industries located in the North[35]. The Tomahawk diagram for two industries is presented in Figure 5[36].

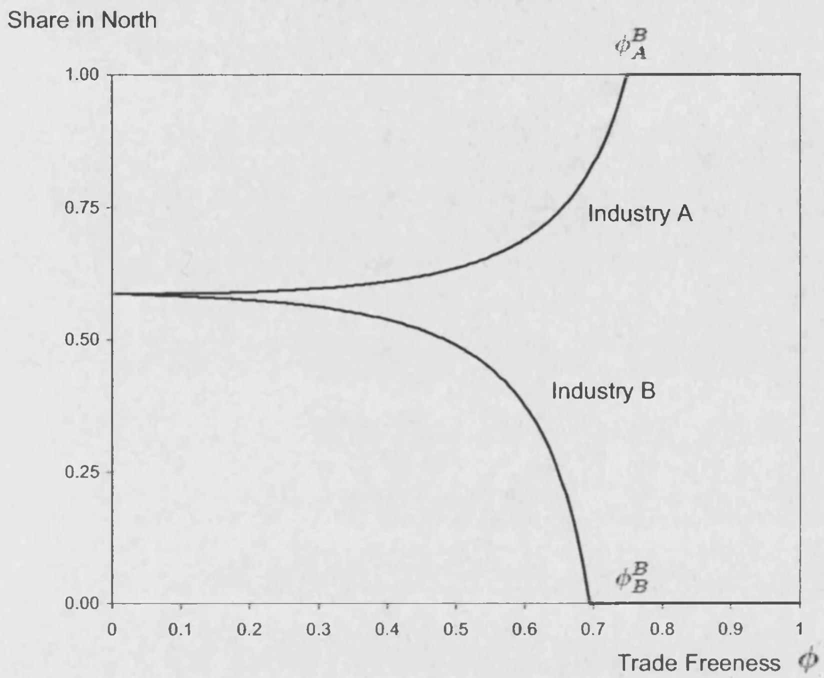
As trade becomes freer (from left to right of the diagram), the breakpoints are reached[37]. At ϕ_A^B the break point of industry A, all firms in this industry are located in the North. At ϕ_B^B the break point of industry B, all firms in

[35]For industry A, North's share is defined as $\frac{n_A}{n_A+n_A^*}$ where n_A and n_A^* are the number of firms in equilibrium (for zero profits) for the North and South respectively. Similarly for industry B, North's share is defined as $\frac{n_B}{n_B+n_B^*}$.

[36]Note that the shares under autarky are not symmetric since North and South do not have the same productivity distribution.

[37]The break and sustained point are the same for a FC model.

Figure 5: Tomahawk Diagram with Industries of Two Different Sunk Costs



the industry are located in the South. Again, it is important to emphasize here that North and South will have the same expenditures for each industry. The implication from the analysis is that as trade becomes freer, industries with low sunk costs will migrate to the South while industries with high sunk costs will migrate to the North. The different profiles of the productivity distributions results in different types of specialisation.

3.7 Conclusion

By synthesising a variant of a New Economic Geography model with recent research into the effects of trade equilibrium under firm heterogeneity, this paper shows that it is possible to rationalise the highly asymmetric allocation

Introducing firm heterogeneity allows the differences between North and South to be modelled by way of firm-level differences rather than through the aggregate production function. With a slight improvement in the North's productivity distribution (first degree stochastic dominance), this paper demonstrates that it is possible to explain the high concentration of firms (and capital) to the North, even though returns to factors of production and expenditures are completely identical between the two regions. The Lucas paradox disappears as a result.

The second key result of the paper demonstrates how the presence of sunk costs in a two-stage entry process can resolve the paradox of risk. 'Hi-tech' or high sunk cost industries tend to locate in the less risky North because it offers them a greater probability of successful entry relative to the South. For 'low-tech' industries with low sunk costs, the North is less attractive since the increase in the probability of entry is offset by the potential of higher post-entry productivity in the South. Capital flows in both directions can be rationalised depending on the level of sunk costs. In a setup with two differentiated sectors, it is possible to show how the high sunk cost industry concentrates in the North and the low sunk cost industry concentrates in the South as trade becomes freer. This result is easily generalised to a multi-industry framework, where the less risky North enjoys a comparative advantage in high sunk cost industries while the South has a comparative advantage in low sunk cost ones. Greater trade liberalisation will lead to both regions specialising in a different set of industries.

The paper also shows how the level of capital flows also depend crucially on the level of trade costs. If trade costs are high, capital will to a large extent be distributed according to expenditure shares. With low trade costs, 'low-tech' industries will locate in the South. This can then explain some stylised differences in the flow of capital to different developing economies. Developing countries with lower trade restrictions will receive more capital particularly from 'low-tech' industries.

3.8 Appendix B

3.8.1 Calibration of Numerical Simulation

Parameter values are referenced to various research where possible. The list of parameters is given in the table below.

Table 9: Parameters and References

Parameters	Value	Remarks
Preferences		
σ	3.8	Referenced to Bernard et al (2003), Ghironi and Melitz (2004) and BRS (2007) estimate of 3.8.
μ	0.5	Arbitrary, no effect on firm aggregates or distribution of firms between the locations, so long as it is small enough such that agriculture continues to exist in both economies.
Endowment		
K_N L_N K_S L_S	1,000,000	Endowments are kept large relative to the fixed cost in order to have an arbitrarily large number of firms in equilibrium. Endowments are symmetric between North and South except for one set of solutions where the home market effect is modelled by increasing North's endowment by 20 per cent.
Pareto Distribution		
$\bar{\varphi}$	0.2	The baseline support is referenced to BRS (2007). However, in the various sets of simulations, the support is varied.
k	3.6	The baseline shape is referenced to BRS (2007). However, in the various sets of simulations, the shape is varied.
Technology		
A	1	Aggregate productivity is normalised to unity for convenience.
α	0.3	This is the capital share in the production of the composite input. Its effect is only on the wage-rental ratio, and has no effect on distribution of firms.

3.8.2 Deriving Total Resource Cost

This subsection proceeds to solve for the total variable production cost in order to pin down the input requirements for the manufacturing sector [see equations (37) and (38)].

Consider a standard total variable production cost (TC) function. This is the integration of the resources used by each firm $\frac{q(\varphi)}{\varphi}$ (marginal cost $\frac{1}{\varphi}$ multiplied by quantity $q(\varphi)$) over the entire distribution of active firms above the cutoff φ^*

$$TC = \int_{\varphi^*}^{\infty} n\frac{q(\varphi)}{\varphi}\frac{g(\varphi)}{1-G(\varphi^*)}d\varphi = nq(\tilde{\varphi})\int_{\varphi^*}^{\infty}\frac{1}{\varphi}\left(\frac{\varphi}{\tilde{\varphi}}\right)^{\sigma}\frac{g(\varphi)}{1-G(\varphi^*)}d\varphi$$

The second equality makes use of the property that $q(\varphi) = q(\tilde{\varphi})\left(\frac{\varphi}{\tilde{\varphi}}\right)^{\sigma}$. With the pareto distribution and the definition of $q(\varphi)$, the above equation can then be simplified to

$$TC = nq(\tilde{\varphi})\frac{k\varphi^{*k}}{\tilde{\varphi}^{\sigma}}\int_{\varphi^*}^{\infty}\varphi^{\sigma-k-2}d\varphi = n\left[\frac{k}{k+1-\sigma}\right]\frac{q(\varphi^*)}{\varphi^*}$$

Total production cost is a $n\left[\frac{k}{k+1-\sigma}\right]$ factor of the variable production cost of the marginal firm $\frac{q(\varphi^*)}{\varphi^*}$.

Consider $\frac{q(\varphi^*)}{\varphi^*}$. Multiplying the numerator and denominator by $p(\varphi^*)$ will give $\frac{q(\varphi^*)}{\varphi^*} = \frac{p(\varphi^*)q(\varphi^*)}{p(\varphi^*)\varphi^*} = \frac{r(\varphi^*)}{p(\varphi^*)\varphi^*}$. Since the marginal firm's revenue $r(\varphi^*)$ must cover σf in equilibrium, and its optimal price is $p(\varphi^*) = \left(\frac{\sigma}{\sigma-1}\right)\frac{1}{\varphi^*}$, it is possible to simplify the equation further to $\frac{q(\varphi^*)}{\varphi^*} = (\sigma-1)f$. This allows the total cost equation to be written as

$$TC = n\left[\frac{k}{k+1-\sigma}\right](\sigma-1)f \tag{B1}$$

Similarly, the total cost to the exporters can be written as

$$TC_X = \tilde{p}_X.n\left[\frac{k}{k+1-\sigma}\right](\sigma-1)\frac{f_X}{\tau}$$

These expressions are then used in equations (37) and (38).

3.8.3 Characterising the Break Point

In FC models, the break points and sustain points are the same since there are no agglomeration effects - whether the initial condition is one of symmetry or asymmetry does not change the outcome. One can begin to solve for the break point by using equation (32). By writing out the aggregate price aggregates explicitly

$$\frac{\varphi_N^{*\sigma-1}}{n\tilde{\varphi}_N^{\sigma-1}+\tilde{p}_X^* n^*\phi\tilde{\varphi}_{SX}^{\sigma-1}} = \frac{\varphi_S^{*\sigma-1}}{\tilde{p}_X n\phi\tilde{\varphi}_{NX}^{\sigma-1}+n^*\tilde{\varphi}_S^{\sigma-1}}$$

Dividing the numerator and denominator of the LHS by $\varphi_N^{*\sigma-1}$ and the right hand side by $\varphi_S^{*\sigma-1}$, one can further simplify the equation to

$$\frac{1}{n+\tilde{p}_X^* n^*\phi Z^{\sigma-1}} = \frac{1}{\tilde{p}_X n\phi Z^{\sigma-1}+n^*}$$

or

$$\tilde{p}_X n\phi Z^{\sigma-1}+n^* = n+\tilde{p}_X^* n^*\phi Z^{\sigma-1}$$

This equation gives the relationship between n and n^*. When all firms are concentrated in the North, $n^* = 0$. Hence

$$\tilde{p}_X \phi Z^{\sigma-1} = 1$$

Recall that $\tilde{p}_X = \left(\frac{\varphi_N^*}{\varphi_{NX}^*}\right)^k = \left(\frac{\varphi_N^*}{Z\varphi_S^*}\right)^k = \left(\frac{\nu}{Z}\right)^k$ giving

$$\left[\frac{\psi^k \tau^k \eta^{\frac{k}{\sigma-1}} - \eta}{\tau^k \eta^{\frac{k}{\sigma-1}} - \psi^k \eta}\right] \eta^{\frac{1}{k}} = \tau \eta^{\frac{1}{\sigma-1}} \tag{47}$$

which provides the implicit solution to the break point - defined as the smallest level of τ that satisfy the above equation. For simplicity, one can assume $f_X = f$ [such as in Falvey, Greenaway and Yu (2006)] or $\eta = 1$. The above equation reduces to

$$\psi^k \tau^k - 1 = \tau \left(\tau^k - \psi^k\right)$$

The bigger advantage the North is given (higher ψ), the higher the τ that can satisfy this condition.

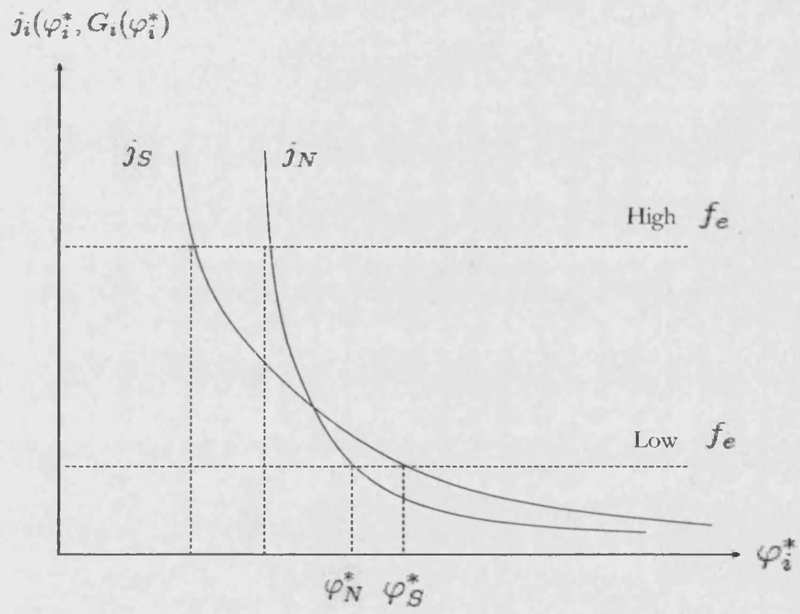
3.8.4 Equilibrium Conditions with Mean Preserving Spread for South

In principle, one can solve for the full equilibrium, including the break/sustain point, in the same manner as when the North is given a productivity advantage. However, because the South's productivity distribution no longer has the same shape as the North $k_N \neq k_S$, it is also not possible for terms to cancel out to arrive at the simple relationship. More importantly, depending on the level of f_e, it is possible that $\varphi_N^* > \varphi_S^*$ (high f_e) or $\varphi_N^* < \varphi_S^*$ (low

f_e) [see Figure 6], making it difficult to generalise the marginal firm condition in equation (32) to an explicit relationship between n and n^*, as in equation (??). Instead, this subsection proceeds to provide some comparative static analytical results, while the numerical results are presented in the main text.

If the North has a higher support but a narrow distribution such that the unconditional means are the same, the equilibrium can be depicted by Figure 6.

Figure 6: Effects of Mean-Preserving Spread of Productivity Distribution



In the case of low f_e, the average productivity of Southern firms is higher than the North since the cutoff is higher there ($\varphi_S^* > \varphi_N^*$) and that the probability mass right of φ_S^* is thicker [see Figure 6]. With low f_e, more firms will have to locate in the South to equalise ex-ante profits between the two locations. Conversely if f_e is high enough, South's cutoff φ_S^* will be low enough relative to φ_N^* such that even the fatter tails cannot compensate. In that case, more firms will locate to the North.

Effects of Increasing Sunk Cost Ignoring the differences between North and South for the moment. Consider only the marginal impact of an increase in the sunk cost f_e. From the ex-ante free entry condition

$$\left(\frac{\bar{\varphi}}{\varphi^*}\right)^k \left(\frac{\sigma-1}{k+1-\sigma}\right)\left\{f + \left[\frac{1}{\tau}\left(\frac{f}{f_X}\right)^{\frac{1}{\sigma-1}}\right]^k f_X\right\} = f_e$$

The mean of a pareto distribution is given as $c = \frac{k\bar{\varphi}}{k-1}$. To keep the mean constant at c while allowing k to vary, the minimum support has to be different. The minimum support can be written as

$$\bar{\varphi} = \frac{c(k-1)}{k}$$

This can be substituted into the previous equation to give

$$\varphi^{*-k}\left[\frac{c(k-1)}{k}\right]^k \left(\frac{\sigma-1}{k+1-\sigma}\right)\left\{f + \left[\frac{1}{\tau}\left(\frac{f}{f_X}\right)^{\frac{1}{\sigma-1}}\right]^k f_X\right\} = f_e \qquad \text{(B4)}$$

Partially differentiating φ^* with respect to f_e gives

$$\begin{aligned}\frac{\partial\varphi^*}{\partial f_e} &= \frac{-\varphi^{*k+1}}{k\left[\frac{c(k-1)}{k}\right]^k \left(\frac{\sigma-1}{k+1-\sigma}\right)\left\{f + \left[\frac{1}{\tau}\left(\frac{f}{f_X}\right)^{\frac{1}{\sigma-1}}\right]^k f_X\right\}} \\ &= \frac{-\varphi^*}{k\varphi^{*-k}\left[\frac{c(k-1)}{k}\right]^k \left(\frac{\sigma-1}{k+1-\sigma}\right)\left\{f + \left[\frac{1}{\tau}\left(\frac{f}{f_X}\right)^{\frac{1}{\sigma-1}}\right]^k f_X\right\}} \qquad \text{(B5)}\end{aligned}$$

In equilibrium, equation (B4) will always hold (envelope condition). This allows equation (B5) to be simplified to

$$\frac{\partial \varphi^*}{\partial f_e} = \frac{-\varphi^*}{k f_e} \tag{B6}$$

This result shows that an increase in f_e always reduces the cutoffs φ^* - this is a standard result. But what are the second order effects when one specifically considers the pareto distribution? Equation (B6) shows that $\frac{\partial \varphi^*}{\partial f_e}$ is more negative at lower level of k (higher variance). The cutoff therefore falls relatively more quickly for the location with the lower k.

The probability of entry $\tilde{p}$ is given as

$$\tilde{p} = \left[\frac{c(k-1)}{k\varphi^*}\right]^k$$

The effect of increase in f_e on entry probability can be found by the partial derivative

$$\begin{aligned} \frac{\partial \tilde{p}}{\partial f_e} &= \frac{\partial \tilde{p}}{\partial \varphi^*}\frac{\partial \varphi^*}{\partial f_e} \\ &= \left[\frac{c(k-1)}{k}\right]^k \frac{\varphi^{*-k}}{f_e} \end{aligned} \tag{B7}$$

Since $\left[\frac{c(k-1)}{k}\right]^k$ is increasing in k, the increase in f_e therefore increases the probability of entry relatively more quickly for a location with higher k (lower variance).

Firstly, equation (B6) says that with an increase in f_e, the cutoff φ^* falls relatively faster for a location with higher variance (which is the South in the context of the discussion). Since average productivity is a function of the cutoff only, this implies that the average productivity falls relatively quickly in the South as well. Secondly, equation (B7) says that the probability of entry $\tilde{p}$ is higher when f_e is higher, but this entry probability increases relatively slower for the location with the higher variance (South). This implies that as f_e rises, the average productivity and probability of entry must rise relatively less in the location with the higher variance (South). As the sunk cost f_e increases, the ex-ante profit of the South falls relative to the North and more firms will have to locate in the North to restore the equilibrium.

4 Evolution of Locations, Specialisation and Returns to Factors With Two Waves of Globalisation

4.1 Introduction

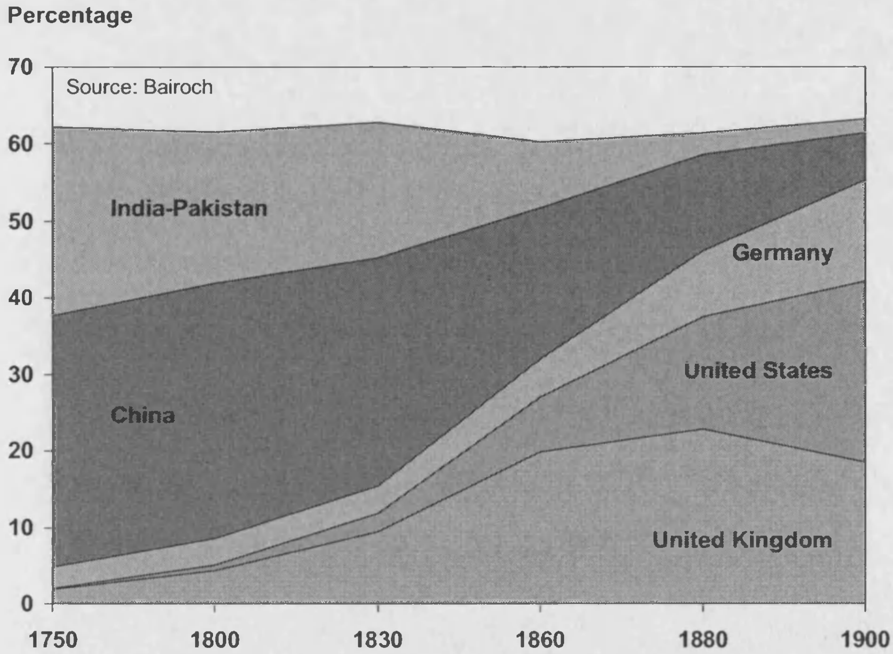
A Historical and Contemporary Overview In 1750, before the Industrial Revolution, China and India produced almost 60 per cent of the world's manufactured goods. Yet, at the turn of the 20th century, the UK was considered the workshop of the world as it produced and exported a huge range of manufactured goods. China and India's share of world manufacturing became minuscule. Manufacturing was also highly concentrated. Three countries - UK, Germany and the United States - accounted for more than 50 per cent of global manufacturing output in 1900 [Figure 7][38].

Today, manufacturing is but a small share of the UK's GDP, falling from 30 per cent in 1973 to 16 per cent in 2003[39]. In its place, a thriving service sector has emerged as epitomised by the city of London. UK exports services to the rest of the world in exchange for manufactured goods. Its per capita service exports are the highest among the G-7 major industrialised nations. In 2006, its service trade recorded a US$53 billion surplus while merchandise trade recorded a deficit of US$143 billion, a persistent pattern of exporting services and importing manufactured goods [Figure 8]. If one considers only two broad categories of production - goods and services - and breaks down the current account as such, a clear pattern of specialisation amongst major developed and emerging economies has also emerged [See Appendix C:4.7.1].

[38] Paul Bairoch, "International Industrialization Levels from 1750 to 1980"

[39] Source: Confederation of British Industries (CBI).

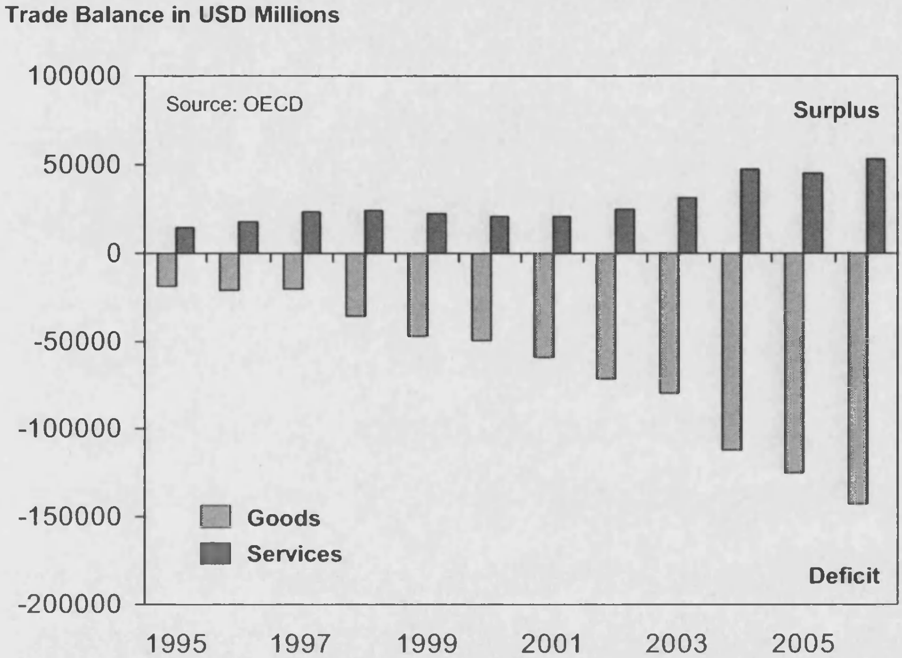
Figure 7: Share of World Manufacturing 1750 - 1900



English speaking countries like the UK, United States, Australia, and even India are increasingly specialising in services and exporting them on a net basis.

On the other side of the world, a different process is taking place. East Asian economies, including China, have been rapidly industrialising as they become more integrated with the global economy. Within the last three decades, China has transformed itself from an agrarian economy to become one of the largest exporters of manufactured goods to the developed countries. There is a great concentration of manufacturing activities in the cities along China's eastern coast and in East Asia more generally. Today, China, Japan and Korea are some of the largest net exporters of manufactured goods, alongside Germany and Brazil [See Appendix:4.7.1]. The evolution of industrial struc-

Figure 8: UK Current Account Breakdown into Goods and Services



ture and the general equilibrium implications for factor prices are important research questions.

Two major themes emerge here. Firstly, some developed economies leading the globalisation process are also witnessing the deindustrialisation and offshoring of manufacturing activities, popularly known as "The Great Sucking Sound". The Anglo-Saxon economies have become some of the largest net importers of manufactured goods, giving rise to the fear of industrial hollowing out and job losses. As Richard E. Baldwin and Philippe Martin (1999) note, "The annual rate of de-industrialisation [of OECD countries] jumped sharply as globalisation picked up pace in the 1980s"[40]. On a smaller spatial scale, the loss of manufacturing for New York and London for example has been well

[40] David Kucera and William Milberg (2002) estimate that trade expansion between OCED and non-OCED countries has resulted in the loss of 3.4 million manufacturing jobs in 10 OECD countries between 1978 and 1990.

documented [Peter Gripaios (1977); Robert Dennis (1978); Frank P. Romo and Michael Schwartz (1995)].

Secondly, there has been a marked increase in inequality as a result of the stagnation of blue-collar wages [John Bound and George Johnson (1992, 1995); Baldwin and Martin (1999)]. More recently, even the offshoring of some white-collar jobs has become a source of concern. It has led some economists to point out the potential causal link between offshoring and wage stagnation [Robert C. Feenstra and Gordon H. Hanson (1999)]. The link between trade, wages and income inequality is still contentious since many economists find skill-biased technological change to be the more plausible explanation of blue-collar wage stagnation. However, the hypothesis that globalisation has resulted in greater income inequality is still a source of keen academic and policy debate, partly because of the contemporaneous nature of de-industrialisation, offshoring, and blue-collar wage stagnation of developed economies.

The key questions remain: Why has so much manufacturing been offshored to developing countries in recent years? Why has the recent wave of globalisation been marked with blue-collar wage stagnation in developed countries and rising within nation inequality [Baldwin and Martin (1999)]? Can the "Shifts in Economic Geography?" [Anthony J. Venables (2006)] be the missing links that explain the contemporaneous occurrence of de-industrialisation and wage inequality?

Two Waves of Globalisation The two waves mentioned in this paper have specific definitions[41]. In this paper, the first wave of globalisation is

[41] The historical stylised facts of the two waves of globalisation are described in detail in Baldwin and Martin (1999).

characterised by goods trade liberalisation. This can be interpreted as the lowering of tariffs for merchandise (manufactured) goods or a fall in shipping or freight costs[42]. The earlier rounds of General Agreement on Tariff and Trade (GATT) contributed to the process of multilateral reduction in import tariffs. In this first globalisation wave, the services sector remained largely protected from foreign competition (i.e., high trade costs) by legal frameworks, regulations, language, or possibly even cultural and social norms. The exact nature of these costs will not be investigated in this paper.

Services here refer to a range of activities such as finance, banking, insurance, consulting, advertising, marketing, legal work that supply to global consumers and businesses [see the global city literature, Saskia Sassen (1991)]. These are the "advanced producer and financial services sectors that serve the command and control requirement of transnational capital" [Neil Brenner (1998)] - which in essence are activities that require a high skill content, as opposed to Balassa-Samuelson type of services like haircuts or plumbing.

In the second wave of globalisation, this paper assumes that services trade costs fall. There are several reasons why one should treat the liberalisation of services trade as a distinct process. The first explanation is due to technology. The sharp fall in telecommunication costs (as opposed to shipping costs for goods), digitisation of information, proliferation of the internet, are more recent phenomena compared to the first wave of globalisation. The fall in communication costs has opened up a whole range of services that can be carried out away from where production of goods takes place or where the

[42] Bairoch (1989) estimates that the 800-km shipment transport cost for iron goods as a percentage of production cost to be 27 per cent in 1830, 21 per cent in 1850, 10 per cent in 1880, and 6 per cent in 1910. Shipping costs therefore had already fallen significantly prior to the First World War.

consumers or downstream firms locate. Similarly, outsourcing of office back-room services would not have been economical without the breakthrough in telecommunication technology. Secondly, many countries have also carried out internal reforms to liberalise their services industries (such as the Big Bang in London's financial services sector), allowing a greater degree of market access by foreign firms. The expansion and deepening of the EU as a single market has also increased market access. Even the greater use of English as a business language can be seen as a reduction of trade cost. These are all relatively recent phenomena compared to the earlier goods trade liberalisation.

Economic Geography in Two Historical Episodes Why might it then be useful to consider trade liberalisation in two episodes? The conclusion of most NEG models is that economic activity can become unevenly spread - allowing a core and a periphery structure to develop [see Anthony J. Venables (2006) for an overview of theory and evidence]. When trade costs are high, production is dispersed in order to serve local markets that cannot otherwise be accessed through trade. When trade costs fall to an intermediate level, agglomeration can result. The salient point about these models is that below a certain level of trade cost (break point), the dispersion of economic activity will not be a stable equilibrium. As Krugman and Venables (1995) show with a horizontal linkages model, a small cost advantage in one location brought about by input-output linkages begets a greater cost advantage by attracting more firms, eventually leading to an outcome characterised by the "Inequality of Nations". Stephen J. Redding and Venables (2004) empirically confirm the importance of market access and sources of supply in explaining variation in per capita income across countries. However, most NEG models do not make any distinction between industries, and are generally silent on which industries actually agglomerate where[43].

Following Masahisa Fujita, Krugman and Venables (1999)[44], this paper assumes two differentiated industries (and one homogeneous industry). To help fix ideas, the two differentiated industries are called manufacturing and services respectively. The first main idea of this paper is that unlike the assumption of many NEG models, trade costs for the two different industries do not fall symmetrically. Rather, this paper assumes that goods trade is liberalised before services trade, thereby allowing manufacturing to agglomerate first.

The agglomeration of manufacturing firms gives rise to an endogenous comparative advantage in the sense that the cost of production (excluding wages) becomes relatively lower. Also, there exist input-output linkages between industries. As manufacturing agglomerates in one location, it also draws in services firms (without services being liberalised) simply because manufacturing firms also buy services. This gives the manufacturing location a head-start in services even in the absence of services liberalisation. When services are finally liberalised, this head start becomes a lock-in cost of production advantage, making services firms agglomerate there as well, but potentially displacing manufacturing (or deindustrialisation) in the process[45].

[43]Except for the distinction between homogeneous agriculture (or the numeraire sector) and the differentiated sector.

[44]Henceforth known as FKV.

[45]A fore-runner to this paper is by Baldwin, Martin and Gianmarco I.P. Ottaviano (2001) - henceforth known as BMO - which shows that intermediate levels of trade costs may cause industries to agglomerate in the North, a result that is consistent with standard NEG predictions. But when the 'cost of trading ideas' falls (that is, greater global spillover of knowledge), it can result in the relative 'deindustrialisation' of the North, which in that exposition is simply the loss of industries (without being precise on what kind of industries

Inequality Within Nations While inequality between nations is the result of early NEG models [Krugman and Venables (1995)], much of the recent debate, controversy about, and opposition to, globalisation is that it creates inequality within nations [Baldwin and Martin (1999)]. On an intuitive level, this is connected to the first theme - namely that manufacturing (including outsourcing of intermediate goods) has relocated from developed western economies to East Asia (first Japan, then the Asian Tigers and finally China) - and how this deindustrialisation has put blue-collar workers under downward wage pressures [Feenstra and Hanson (1999)]. Furthermore, Brenner (1998) notes that social research into global cities has been dominated by some inter-related themes, including "deindustrialisation, . . . , expansion and spatial concentration of financial and producer services industries, labour market-segmentation, . . . , socio-spatial polarisation."

In summary, the contribution of this paper is to make use of what is already a standard NEG model and embed it with a multi-industries and multi-factors setup, in which globalisation influences the changes in industrial structure, eventually leading to both within and between nation inequality.

Limitations NEG models are not without their critics[46]. Firstly, many of the results stem from a specific functional form - namely CES preferences. Secondly, some results such as the home market effect depend on the simplifying assumption that there is a homogeneous good (or sometimes known as agri-

are lost). Unlike BMO (2001), there is no accumulation of capital, no technology spillovers, no learning effects and no 'cost of trading ideas'. Furthermore, this paper shows that the loss of manufacturing is not merely relative, but absolute. The second and distinct wave of globalisation can result in a shift in the endogenously determined comparative advantage - away from manufacturing and to services. As the North gains services firms, it can lose manufacturing firms in absolute terms.

[46]Peter J. Neary (2000) provides an interesting exposition on the shortcoming of various NEG models.

culture) that is competitively produced and costlessly traded to equalise the wage between two locations. Donald R. Davis (1998) shows that the costless trading assumption is not supported empirically, and that the home market effect disappears if one assumes trade costs to be the same for industrial and homogeneous goods. Thirdly, many NEG models make use of the simplifying assumption that consumption of the differentiated good is a small proportion of total income which ensures incomplete specialisation in equilibrium[47]. This assumes away any congestion cost that arises from agglomeration, leading to 'bang-bang' predictions - that is, full agglomeration of all industries in one location or the other - that are not observed in reality. Finally, NEG models are often analytically intractable, particularly those with feedback loops, and are often only solved by numerical methods.

Many of the simplifying assumptions employed in many NEG models will also be used in this paper, and many of the criticisms will also apply. In particular, the model still relies on the agriculture good as the numeraire. It also relies on numerical solutions to illustrate the key points. In defence however, the model that this paper uses is one where input-output linkages drive the agglomeration process. Its conclusions therefore do not depend on the home market effect. In addition, as mentioned in the introduction, one of the objectives of this paper is to capture the displacement effect. This essentially is to move away from the 'bang-bang' predictions of many NEG models, thereby generating more realistic outcomes for industrial locations.

[47] An alternate way of putting this is that the agricultural sector continues to operate in both regions to equalise wages. The monopolistic but atomistic firms in the industrial sector effectively face an infinite supply of labour at the given wage rate.

4.2 Model Setup

4.2.1 Regions

There are two regions in the model, North and South, subscripted by ($j = N, S$). There are three industries - the agricultural industry A acting as the walrasian, costlessly transported, numeraire good, and two differentiated industries - manufacturing and services. Factors of production are immobile between the two regions.

4.2.2 A Hierarchy of Skills and Industries

There are three primary factors in the model - high-skilled (K), semi-skilled (L) and unskilled labour (O). All workers are immobile between countries. There is a hierarchy of jobs that they can perform. High-skilled labour can work in all industries. Semi-skilled skilled workers can only work in manufacturing and agriculture; they are unproductive in services. Unskilled workers can only work in the agricultural industry, and are unproductive if used in the other two sectors. The skill level of a worker is therefore characterised by the range of tasks that he can perform. For example, a PhD can become equally productive as a farmer when he is deployed to a farm, but a farmer has zero productivity when deployed to a university. Succinctly put, a skilled worker is a perfect substitute for all workers with skills below him but an unskilled or semi-skilled worker is not a substitute for the worker type above him.

There are many reasons why this specification is attractive. Consider an alternative formulation where one allows the more productive worker a Ricardian improvement in his productivity. In effect, this specification then becomes a unit of measure issue - that is, effective labour units. If worker K is twice as productive as worker L, he simply becomes two times worker L and draws twice the wage. In a general equilibrium analysis, one just needs the knowledge of effective labour units to pin down the market size. Furthermore, it also does not matter where each industry agglomerates. Within each location, the ratio of wages will just reflect the Ricardian productivity difference in any full employment equilibrium.

However, by reformulating the problem into one that has a hierarchy of jobs, the effective constraint placed on each industry becomes different. The lowest order industry - agriculture - has the highest potential pool of workers. The services industry, which can only use K-type workers, has the smallest potential pool of workers. The manufacturing industry is in between. When agglomeration takes place, different industries will have different effective constraints on their expansion. In NEG parlance, the congestion costs will be different for various industries, thereby resulting in different equilibrium wages where different industries agglomerate[48].

To prevent any Hecksher-Ohlin motivation for trade, this paper assumes that both countries have the same endowments of unskilled, semi-skilled and skilled workers[49]

$$O_N = O_S = \bar{O} \qquad L_N = L_S = \bar{L} \qquad K_N = K_S = \bar{K}$$

[48] The labour market need not be quite so dichotomous. For example, one can allow both manufacturing and services to use both skilled and semi-skilled workers, with services having a higher skill intensity compared to manufacturing. It is possible to create the same qualitative result. However, the choice of endowment will become more important in this alternative setup. For example, in the first wave of globalisation when manufacturing agglomerates, the wages for semi-skilled workers might rise above skilled workers since they are used more intensively, thereby presenting interpretational difficulties. With the setup presented in this paper, this anomaly will never arise since K-type workers are a perfect substitute for L-types but not the other way around. Less skilled workers will never receive more than their more skilled counterparts in any equilibrium configuration.

[49] This will be relaxed later to highlight the effect of endowment on changes in industrial specialisation and wages.

Even though both locations have exactly the same endowments, "comparative advantage" nevertheless arises in equilibrium because of the agglomeration process driven by the existence of input output linkages.

The characterisation of agricultural unskilled workers requires a little more explanation. In standard NEG models, workers are homogeneous. The presence of the agricultural sector in both regions, implied by a small consumption share of industrial goods, makes it easy to characterise the equilibrium. The agricultural sector simply serves as the buffer sector; any workers not used in industrial production are deployed in agriculture. Labour market clearing is achieved with great simplicity. The same mechanism is at work here with some modifications.

Effectively, the labour market works this way. If the supply of skilled labour is greater than the demand in services for it, the excess supply will be added to the manufacturing sector alongside the semi-skilled labour pool. Given the manufacturing sector labour demand for semi-skilled workers, the excess supply of semi-skilled workers will be used in the agricultural sector. Excess labour effectively cascades or overflows downwards. Unskilled labour, together with what is left from manufacturing and services, then becomes the buffer in the same way that the agricultural sector acts as the buffer in many NEG models.

4.2.3 Consumer

Each worker, regardless of type, maximises utility over all types of goods in a simple Quasi-linear utility[50]

$$U = A + \mu_M \ln X_M + \mu_S \ln X_S \tag{48}$$

where $X_M = \left[\int x_{Mi}^{\frac{\sigma-1}{\sigma}} di\right]^{\frac{\sigma}{\sigma-1}}$ and $X_S = \left[\int x_{Si}^{\frac{\sigma-1}{\sigma}} di\right]^{\frac{\sigma}{\sigma-1}}$ are the CES aggregated manufacturing and services good respectively, and μ_M and μ_S measures the intensity of consumption of both sectors. Furthermore, X_S is aforementioned to be advanced producer services, it is reasonable to assume that $\mu_S < \mu_M$ - in other words, a consumer lower direct purchases from this sector[51].

From equation (48), the demands of each consumer for agriculture, manufacturing and services are given as follows

$$D_A = w - \mu_S - \mu_M \qquad D_M = \frac{\mu_M}{P_M} \qquad D_S = \frac{\mu_S}{P_S} \tag{49}$$

where w is the wage, P_M and P_S are the CES aggregated prices. The coefficients are constrained $\mu_S + \mu_M < w$ so that there will be a positive consumption of the A good.

Within each differentiated sector, each consumer's demands still take the

[50]The model works with a Cobb-Douglas utility function as well. In equilibrium, different worker types will be paid different wage rates. With Cobb-Douglas preferences, this implies that expenditures on each class of good is different for each type of worker. The Quasi-linear preferences allow us to ignore these income effect considerations. All workers, regardless of how much they are paid in equilibrium, will demand the same amount of X_M and X_S. Higher income translates to higher demand for the A good. The upshot of this is that there will not be any "home market effect" arising from consumer expenditure for the differentiated industries. All agglomeration is due to forward and backward linkages on the production side.

[51]This assumption has no bearing on the qualitative pattern of results except to note that, if the intensity of demand for services is too high, the equilibrium will show dispersion of the services industry even at low levels of services trade costs since skilled workers are assumed to be evenly spread between the two locations.

usual CES form

$$d_M = \frac{p_M^{-\sigma} \mu_M}{P_M^{1-\sigma}} \qquad d_S = \frac{p_M^{-\sigma} \mu_S}{P_S^{1-\sigma}} \tag{50}$$

In equilibrium, different worker types will potentially receive different wages due to the presence of agglomeration rents. But with the quasi-linear function, there are no income effects for the differentiated sectors. Therefore, following (50), the firm level demands become

$$\bar{d}_M = \frac{p_M^{-\sigma} \bar{\mu}_M}{P_M^{1-\sigma}} \qquad \bar{d}_S = \frac{p_S^{-\sigma} \bar{\mu}_S}{P_S^{1-\sigma}} \tag{51}$$

where $\bar{\mu}$ simply aggregates across the population of consumers (held as a constant throughout), since all of them have the same expenditure on manufacturing and service.

4.2.4 Production Technology

Production in sector A is standard - 1 worker to produce 1 unit of the good. Production in sectors M and S exhibits increasing returns to scale with a fixed cost f and variable cost κ. The cost function is homothetic, both fixed and variable costs use the factors in the same intensity. In addition, there are input-output linkages. For manufacturing, within or intra-industry linkages are captured by α_M and external or inter-industry linkages are captured by β_M. Intra-industry input-output linkages are assumed to be stronger than inter-industry input-output linkages. The labour share is $\gamma_M = 1 - \alpha_M - \beta_M$.

The total cost function for M type firm in country j is therefore

$$C_{M_j} = (F + \kappa x_{Mj})\, w_j^{\gamma_M} P_{M_j}^{\alpha_M} P_{S_j}^{\beta_M} \theta_M \quad (52)$$

Similarly, the cost function for firm S firm in country j becomes

$$C_{S_j} = (F + \kappa x_{Sj})\, w_j^{\gamma_S} P_{S_j}^{\alpha_S} P_{M_j}^{\beta_S} \theta_S$$

where P_j is the CES aggregated price over all differentiated goods, κ is the per unit input requirement, and with θ_M and θ_S simply the constants of cost minimisation[52].

4.2.5 Trade Cost

Both manufacturing and services sectors face iceberg trade cost when exporting goods and services. However, in contrast to most Economic Geography models, this paper has elected to model different trade costs τ_M and τ_S for manufacturing and services respectively. The reason behind this is to allow the globalisation process to occur as two distinct and separate processes for the two industries, roughly in keeping with the historical evidence.

The first wave of globalisation is captured by a decrease in τ_M from some arbitrarily high levels, thus freeing up goods trade. While trade in goods becomes freer, trade in services is still not possible due to prohibitive trade barriers - high communication costs, language, legal, regulatory barriers - all captured by τ_S. The second wave of globalisation captures the effect of trade

[52]It is not necessary to assume that the linkages are symmetric between the two differentiated industries. The input-output table (see Appendix) suggests they are not. However, in the numerical simulations later, the inter and intra-industry linkages are nevertheless assumed to be symmetric as this has no relevance for the qualitative result.

liberalisation in services. Following standard notation, two freeness (phi-ness) of trade indices for the two sectors are defined as phi-M ($\phi_M = \tau_M^{1-\sigma}$) and phi-S ($\phi_S = \tau_S^{1-\sigma}$) respectively.

4.2.6 Unit Cost and Prices

The CES function gives the standard aggregated prices for the M or S sectors, P_M and P_S respectively (and with asterisk for the South). Given the description of the wages above, the unit cost functions can be written as

$$\begin{aligned} c_M &= w_L^{\gamma_M} P_M^{\alpha_M} P_S^{\beta_M} \theta_M \qquad & c_M^* = w_L^{*\gamma_M} P_M^{*\alpha_M} P_S^{*\beta_M} \theta_M \\ c_S &= w_K^{\gamma_S} P_S^{\alpha_S} P_M^{\beta_S} \theta_S \qquad & c_S^* = w_K^{*\gamma_S} P_S^{*\alpha_S} P_M^{*\beta_S} \theta_S \end{aligned} \tag{53}$$

Therefore, the unit cost functions depend not only on the aggregate prices, but also the kind of wages faced by each industry (which may no longer be 1 for every worker everywhere). The wage of semi-skilled and skilled workers w_L and w_K, together with the counterparts in the South w_L^* and w_K^*, will be endogenised in equilibrium. Their values will depend on the patterns of agglomeration that emerge.

4.2.7 Profits

The profit function of each manufacturing firm in the North can be written as

$$\pi_M = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\kappa c_M\right)^{1-\sigma}\left\{\frac{\bar{\mu}_M + \alpha_M R_M + \beta_S R_S}{P_M^{1-\sigma}} + \phi_M\left[\frac{\bar{\mu}_M^* + \alpha_M R_M^* + \beta_S R_S^*}{P_M^{*1-\sigma}}\right]\right\} - c_M F \tag{54}$$

where R_M and R_S are North's aggregate intermediate purchases by the manufacturing and services sector respectively (to be defined later). Hence, the manufacturing firm charges an optimal price of $\frac{\sigma}{\sigma-1}\kappa c_M$, which is a constant mark up over its unit cost. The terms inside the curly brackets provide the total market potential perceived by the firm from home sales (first term) and export sales (second term). The fixed cost is given by $c_M F$. Since the two locations have the same population size, $\bar{\mu}_M = \bar{\mu}_M^*$. Similarly for the services industry, the profit function can be written as

$$\pi_S = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\kappa c_S\right)^{1-\sigma}\left\{\frac{\bar{\mu}_S + \alpha_S R_S + \beta_M R_M}{P_S^{1-\sigma}} + \phi_S\left[\frac{\bar{\mu}_S^* + \alpha_S R_S^* + \beta_M R_M^*}{P_S^{*1-\sigma}}\right]\right\} - c_S F$$

If more manufacturing firms locate to the North, the market potential in the North increases relative to that in the South because of the presence of backward linkages. There will be more firms at the same location that demand its products as intermediates. At the same time, because of the intra-industry linkages, the unit cost in the North will fall relative to that in the South (forward linkages) since the intermediates are no longer subjected to trade costs.

In the South, the profit functions are exactly analogous and are given by

$$\pi_M^* = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\kappa c_M^*\right)^{1-\sigma}\left\{\phi_M\left[\frac{\bar{\mu}_M + \alpha_M R_M + \beta_S R_S}{P_M^{1-\sigma}}\right] + \frac{\bar{\mu}_M^* + \alpha_M R_M^* + \beta_S R_S^*}{P_M^{*1-\sigma}}\right\} - c_M^* F$$

(55)

$$\pi_S^* = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\kappa c_S^*\right)^{1-\sigma}\left\{\phi_S\left[\frac{\bar{\mu}_S + \alpha_S R_S + \beta_M R_M}{P_S^{1-\sigma}}\right] + \frac{\bar{\mu}_S^* + \alpha_S R_S^* + \beta_M R_M^*}{P_S^{*1-\sigma}}\right\} - c_S^* F$$

Some Normalisations To simplify the notations, the paper makes a few convenient normalisations [see Economic Geography & Public Policy by Baldwin et al]. Firstly, $\kappa = \frac{\sigma-1}{\sigma}$ thereby simplifying the cost of each unit to 1. Secondly, $F = \frac{1}{\sigma}$. Take for example a profit function in equation (54). By using these normalisations, the profit function can be written as

$$\pi_M = \frac{1}{\sigma} c_M \left\{ c_M^{-\sigma} \left[\frac{\bar{\mu}_M + \alpha_M R_M + \beta_S R_S}{P_M^{1-\sigma}} + \phi_M \left(\frac{\bar{\mu}_M^* + \alpha_M R_M^* + \beta_S R_S^*}{P_M^{*1-\sigma}} \right) \right] - 1 \right\}$$

where the terms inside the curly brackets represent output given the CES preference. Given these two normalisations, the paper makes use of the standard result that the production scale consistent with zero-profit is

$$\bar{x} = 1$$

This greatly reduces the notations and also simplifies later numerical analysis. Note that in equilibrium, the total input requirement of each firm is given as $(\bar{x}\kappa + f)$, which because of the above normalisation is also equal to 1. Each firm therefore will demand one physical unit of input in equilibrium.

4.2.8 Workers' Types and Equilibrium Wage Rates

The price of good A is chosen as the numeraire. Since unskilled workers are immobile to other industries, the agricultural sector must operate in both countries, thereby pinning down low-skilled wages as $w_O = 1$. If all K-type workers are employed in the S sector, then $w_K \geqslant w_L$ in equilibrium. Otherwise, K-type workers would be better off in the manufacturing sector, and the labour market cannot be in equilibrium. Writing these explicitly

$$\bar{w}_K = \left\{ \begin{array}{ll} w_K & K_N^d = K_N \\ w_L & K_N^d < K_N \end{array} \right\} \tag{56}$$

where $\bar{w}_K$ indicates the equilibrium wage and K_N^d is the conditional demand for skilled labour in the North. This is given by

$$K_N^d = \left(\frac{\gamma_S}{\alpha_S}\right)^{\alpha_S} \left(\frac{\gamma_S}{\beta_S}\right)^{\beta_S} \left(\frac{P_S}{w_K}\right)^{\alpha_S} \left(\frac{P_M}{w_K}\right)^{\beta_S} n_S \tag{57}$$

which is simply the demand per services firm multiplied by the number of services firms in the North n_S since each firm demands 1 unit of input given the normalisation. Equation (57) is therefore the total derived demand for skilled labour given the number of firms operating in that particular location[53].

In equilibrium, it is possible that not all K-type workers will work in services ($K_N - K_N^d > 0$). In that case, those K-type workers not employed in services will choose to work in manufacturing and be offered wage w_L. The wage function of w_L becomes

$$\bar{w}_L = \left\{ \begin{array}{ll} w_L & L_N^d = L_N + (K_N - K_N^d) \\ w_O = 1 & \text{Otherwise} \end{array} \right\} \tag{58}$$

where conditional demand for manufacturing labour is given as

$$L_N^d = \left(\frac{\gamma_M}{\alpha_M}\right)^{\alpha_M} \left(\frac{\gamma_M}{\beta_M}\right)^{\beta_M} \left(\frac{P_M}{w_L}\right)^{\alpha_M} \left(\frac{P_S}{w_L}\right)^{\beta_M} n_M \tag{59}$$

[53] Off equilibrium, if conditional demand is larger than supply of factor $K_N^d > K_N$, wage w_K has to increase until the equilibrium value $\bar{w}_K$ solves for $K_N^d = K_N$. If conditional demand is not larger than the supply at that location, $\bar{w}_K$ must fall to that of the next skill tier w_L.

and n_M is the number of manufacturing firms in the North[54]. The same set of equations for w_K^* and w_L^* can also be written for the South analogously, providing a total of four wage equations.

The above equations also give a sense on why an analytical solution might prove difficult. The wage functions are non-differentiable as they are discontinuous at the point at which the supply constraint binds. In contrast, the wage rate is determined by the price of the agricultural good in an equilibrium characterised by incomplete specialisation. As can be seen from the above equations, wages are a function of local conditional demands (which depend on the number of firms at each location), which is also a function of wages. The mapping of wages onto firms and then back onto wages makes it difficult to solve this problem analytically[55].

4.2.9 Definition of Intermediate Revenues and Total Wage

Intermediate revenues are given in a standard way, simply the total cost of all firms. The intermediate revenues for North and South manufacturing are

$$R_M = n_M c_M(\kappa\bar{x} + F) \qquad R_M^* = n_M^* c_M^*(\kappa\bar{x} + F) \tag{60}$$

Similarly, the intermediate revenues for North and South services are

$$R_S = n_S c_S(\kappa\bar{x} + F) \qquad R_S^* = n_S^* c_S^*(\kappa\bar{x} + F)$$

[54] Again, in an off equilibrium situation where $L_N^d \geqslant L_N + (K_N - K_N^d)$, then w_L must increase until the equilibrium value $\bar{w}_L$ solves for $L_N^d = L_N + (K_N - K_N^d)$. If $L_N^d < L_N + (K_N - K_N^d)$, then w_L will take the value of 1 since some semi-skilled workers will work in the agricultural sector.

[55] The wage functions are also non-differentiable.

Total incomes (which are equal to expenditures) are given by

$$E = w_K K_N + w_L L_N + O_N \qquad E^* = w_K^* K_S + w_L^* L_S + O_S \quad (61)$$

With these, agriculture demands are given by

$$A = E - \bar{\mu}_M - \bar{\mu}_S \qquad A^* = E^* - \bar{\mu}_M - \bar{\mu}_S$$

With the quasi-linear preferences, only the demand for agriculture is affected by changes in wage incomes.

4.3 Solving for Equilibrium

4.3.1 Equilibrium Conditions

Typically in a NEG model, the number of firms is fixed in the short-run while other endogenous variables adjust. In the long run, all the conditions of the short-run equilibrium are met while allowing the free entry conditions (firms' entry and exit) to be satisfied. But since this paper is concerned with the long-run evolution of industrial locations and factor returns over two episodes of globalisation, it makes no distinction between the short and long-run solution. Instead, all variables are allowed to adjust towards the long-run equilibrium from the onset for every level of given trade costs. The equilibrium is characterised by a vector of eight endogenous variables $\{n_M$, n_M^*, n_S, n_S^*, w_K, w_L, w_K^*, $w_L^*\}$ that are pinned down by the four zero-profit conditions and four wage equations such that:

(a) all firms make zero profits given entry, exit and relocation and;

(b) goods markets clear and;

(c) there is no excess demand or supply for skilled, semi-skilled or unskilled labour given the equilibrium wages in both locations.

4.3.2 Adjustment Process

Given a specific level of trade costs, the numerical solutions begin by imposing a symmetric (arbitrary) number of firms to both locations. A small positive shock ($\varepsilon_F = 0.01$) to the number of manufacturing firms is given to one location, and a small negative shock ($-\varepsilon_F$) is applied to another, to break the initial symmetry.

Step 1 Firms are allowed to enter, exit or relocate. If both North and South firms have positive profits given the number of existing firms, a small increment ($+\varepsilon_F$) in the number of firms is further applied to both locations - this is entry. If both North and South firms continue to have negative profits, a small decrease ($-\varepsilon_F$) in the number of firms is applied due to exit. If the North has higher profits than the South, firms migrate northwards ($+\varepsilon_F$ for the North and $-\varepsilon_F$ for the South) for relocation. Vice versa.

Step 2 Local labour markets adjust. Given the interim number of firms in both North and South after the step 1 adjustments, the demand for labour for the industry for each location can be derived using the cost minimising demand function [see equations (57) and (59)]. If demand for skilled labour from services exceeds supply at a particular location, the local skilled wages will move upwards in a small increment ($\varepsilon_W = 0.0001$)[56]. If not, the excess skilled labour is added to the local pool of semi-skilled labour or downward cascading.

[56]The adjustment parameters ε_F and ε_W represent a search increment of a magnitude that is one-hundredth of a per cent of the underlying variables.

Again, the demand for labour from manufacturing is checked against this pool. If demand exceeds supply, even the semi-skilled wage moves upwards (also bumping up skilled wages in the process since skilled wages cannot be lower than semi-skilled labour in equilibrium). Otherwise, any excess overflows into the unskilled labour pool too, which then serves the numeraire sector agriculture.

Iteration Steps 1 and 2 are iterated until the equilibrium conditions stated are met. This will then give the long-run number of firms and wages in each location.

4.3.3 Model Parameters

The key parameters in this model are the intra and inter-industry linkages as they determine the strength of the agglomeration forces as well as the wages in equilibrium. In a two-sector setup, it is necessary that the intra-industry linkages are stronger than inter-industry ones for agglomeration to take place. Otherwise, both industries will tend to disperse rather than agglomerate since the benefits of co-location with another industry is greater than each industry locating in one location [see FKV for further exposition]. The input-output table provides strong evidence that intra-industry linkages are indeed stronger [see Appendix C:4.7.2][57].

[57] In the numerical simulations, this paper assumes the intra- and inter-industry linkages of both sectors to be symmetric. Though the IO table suggests that they are slightly different, it does not change the results qualitatively, so long as intra-industry linkages are stronger.

4.4 Results

4.4.1 Autarky Results

Economic activity is completely dispersed in the autarky equilibrium. All labour types have the same wage, equal to the numeraire. This is due to the fact that the paper has chosen factor endowment such that factor supplies of skilled and semi-skilled workers exceed the demands for them in the absence of agglomeration. Since prices are also the same in both locations given the symmetry of the regions, real wages are also equal. There is no inequality between or within nations.

4.4.2 First Wave of Globalisation

In the first wave of globalisation, trade costs in services are kept arbitrarily high such that $\phi_S = 0$ (no services trade is possible). In the numerical solutions, the trade cost for manufacturing is gradually lowered (increase in phi-M). Two sets of equilibrium paths are possible - one with greater a concentration of both manufacturing and services in the North (shares greater than 0.5) and the other with a greater concentration in the South. This can be seen in Figure 9, where the Y-axis shows the share of firms for each industry located in the North plotted against phi-M (X-axis).

Wage-sustain and Wage-break Points The definitions of break and sustain points are slightly different from traditional NEG models since the parameters and endowments are chosen so agglomeration will result in the rise of wages due to the limited endowments. In a symmetric equilibrium, both locations have exactly the same number of services and manufacturing firms, and wages everywhere for all types of workers will be the numeraire wage ($w_K = w_K^* = w_L = w_L^* = 1$).

In order to make a distinction between the break/sustain point as used in the standard NEG literature, this paper introduces the "wage-break or wage-sustain points"[58]. The wage-break point is defined as the level of trade cost (which can be equivalently expressed in terms of trade freeness) that allows wage symmetry between the types of labour within a location to be broken. The wage-sustain point is the level of trade cost that allows a wage differential between the types of labour to be sustained in a location.

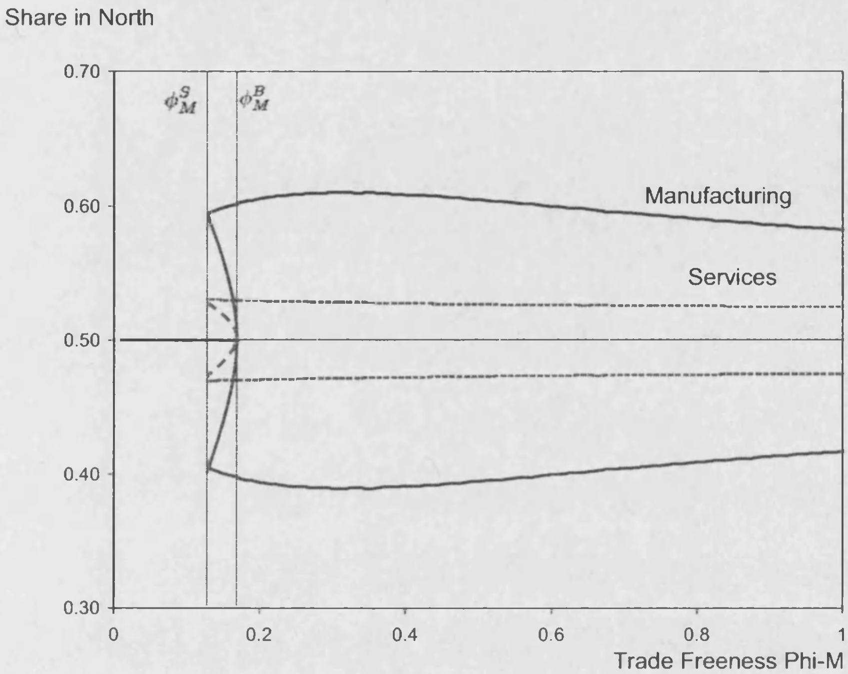
For example, the manufacturing wage-break point is the level of trade freeness ϕ_M^B which allows the symmetric distribution of manufacturing to break in favour of one location, resulting in a premium on manufacturing wages in that location. Similarly, the manufacturing "wage-sustain point" is the level of trade freeness ϕ_M^S that allows manufacturing wages to be sustained at a higher level, with the maximum level of agglomeration as the initial condition.

Tomahawk Diagram The tomahawk diagram plots the shares of firms of the North (South's shares are simply the complement) on the Y-axis against the level of trade freedom phi-M on the X-axis. The firm share diagram is a 'multiple pitchfork' as there is more than one industry. The wage-break and wage-sustained points can be clearly seen in the Figure 9[59].

[58] In the standard NEG literature, the break point is the value of trade costs at which the symmetric equilibrium with equal shares of industry between two locations becomes unstable. The sustain point is the value of trade costs at which the asymmetric equilibrium (with agglomeration) can be sustained.

[59] As standard in NEG models, the sustain point comes before the break point since the value of trade costs at which the symmetric equilibrium becomes unstable is lower (or phi-ness higher) than the value of trade costs at which the asymmetric equilibrium (agglomeration) can be sustained. There is a region of overlap between the two (see ϕ_M^B and ϕ_M^S in Figure 9) where both the agglomeration or symmetric equilibrium are stable, depending on the initial conditions. Because of the possibility of multiple equilibria in this region of overlap,

Figure 9: Tomahawk Diagram (First Wave)



The outer pitchfork lines (in solid blue) show the equilibrium shares of manufacturing while the inner pitchfork shows the equilibrium shares of services (in dotted red). Though there is no services trade, its shares are influenced by the shares of manufacturing as a result of the inter-industry linkages, as shown in the inner pitchforks. Since supply of skilled and semi-skilled labour is finite and wages increase when these constraints are reached, the pitchfork lines represent the maximum shares that do not reach 1 or 0. In other words, the equilibrium does not exhibit a 'bang-bang' outcome.

The level of endowments relative to the size of consumption will determine how much agglomeration can take place. For example, given a constant con-

expectations become important since some firms may shift location in anticipation that this will change the initial conditions enough to trigger agglomeration. To keep exposition simple, this paper ignores the role of expectation.

sumption intensity in the preferences, a symmetric increase in the endowment of K-type labour in both the North and South will widen the pitchforks by allowing more agglomeration to take place before resource constraints start to bind.

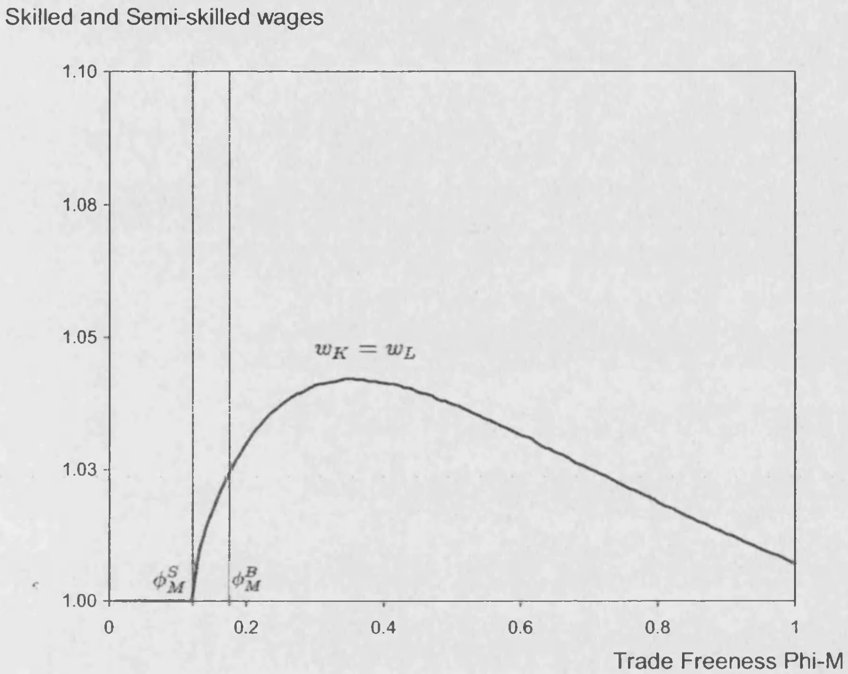
Agglomeration Rents When the labour resource limits are reached, labour begins to earn agglomeration rents. For example, when manufacturing agglomerates in one location and exhausts the pool of skilled and semi-skilled labour, entry firms bid up the equilibrium wages according to equations (56) and (58). The evolution of wages is given in Figure 10. As a result of manufacturing agglomeration, skilled and semi-skilled workers begin to command a wage premium over unskilled workers (whose wage is set to 1). The standard hump-shape feature of the equilibrium wage path is that inequality is highest at intermediate levels of merchandise trade costs - rising quickly after the wage-sustain point ϕ_M^S and falling gradually as trade cost falls further.

From the wage break point ϕ_M^B, there is a small discrete jump in the wages as the symmetric equilibrium is broken. In this first wave, the change in wage structure is driven by the agglomeration of manufacturing. As manufacturing uses both K and L types, the agglomeration rent is shared between these two types of workers. This also implies that services demand for skilled workers does not exhaust the supply of all skilled workers, and some of them are employed in the manufacturing sector in equilibrium.

4.4.3 Second Wave of Globalisation

Goods trade costs are held at a constant level ($\phi_M = 0.7$ or $\tau_M = 1.10$). Given this level of ϕ_M, the paper solves for the long-run equilibrium when

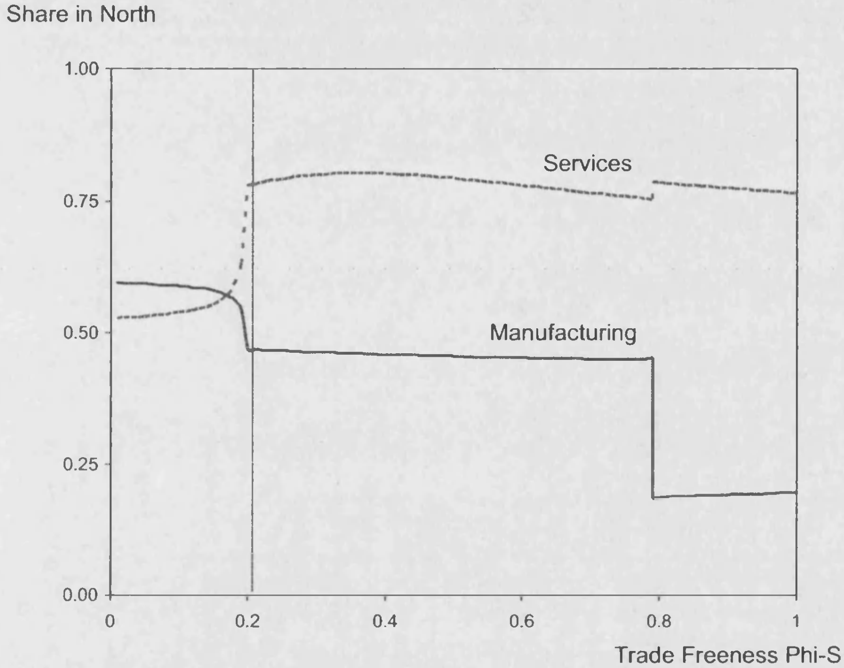
Figure 10: North's Equilibrium Wages (First Wave)



services trade becomes free (from $\phi_S = 0$ to 1). The services "wage-break point" is the level of services trade cost, below which services reaches maximum agglomeration and exhausts the sector specific factor. In other words, the services wage-break point is reached when K type wages become higher than L type wages ($w_K > w_L$ or $w_K^* > w_L^*$).

In order to reduce the number of possible equilibrium development paths and to simplify the exposition, the paper shows only one set of equilibrium paths for the second wave of globalisation - one that assumes that the North has a larger share of manufacturing in the first wave. This is broadly consistent with the historical pattern since the Industrial Revolution. The fact that trade costs fall sequentially (goods first then services) becomes important here. The key thing to note is that even though services is not liberalised in the first wave, the North has an initial advantage in services. From an arbitrarily high level of trade costs, services trade becomes freer (moving right on the X-axis in Figure 11).

Figure 11: Tomahawk Diagram (Second Wave) North



As services trade is freed, services sector firms begin to agglomerate in the North quickly as a result of the initial advantage (see Figure 11). As skilled workers begin to migrate away from manufacturing and into services, there is a contraction of manufacturing activities in the North, relocated to the South. The wage-break point occurs at around $\phi_S = 0.2$ (this can be related to Figure 12, at the same ϕ_S where w_K becomes higher than w_L). There is location hysteresis, but only for the services sector. Since this is a model with horizontal linkages, the relocation of manufacturing activities to the South effectively becomes offshoring since firms in the North will use a greater share of intermediates produced in the South (pseudo-offshoring).

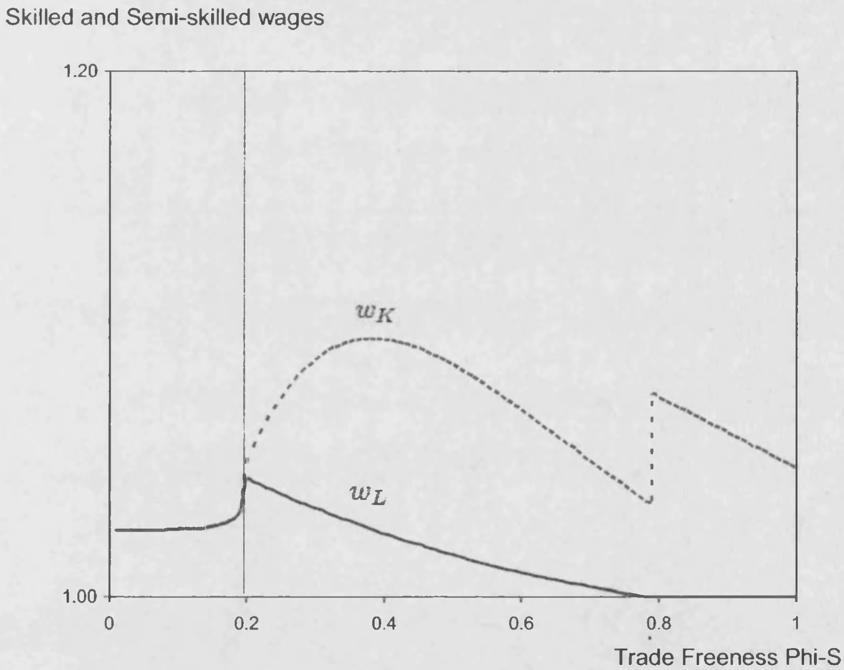
When services trade becomes very liberalised at around $\phi_S = 0.8$ (Figure 11), the North experiences a precipitous loss of manufacturing or large scale deindustrialisation. This discontinuity occurs because of the following reason. When services trade is initially liberalised, the presence of a large services sector agglomerating in the North offers manufacturing firms lower cost of production, since manufacturing firms also demand services intermediates. The inter-industry forward linkages induce a sizeable number of manufacturers to maintain their presence in the North at intermediate levels of services trade liberalisation (from ϕ_S values of 0.2 to 0.8). However, further liberalisation of services can bring the equilibrium into a tipping point when even the South can access services intermediates from the North cheaply. North's advantage in cheaper services intermediates becomes outweighed by the cheaper labour in the South (which is $w_L^* = 1$). As a manufacturing firm relocates to the South to take advantage of the cheaper cost of labour there, it further reinforces the attraction of the South through intra-industry input-output linkages. Because of this process of cumulative causation, more manufacturing firms relocate to the South until equilibrium is restored[60]. The key point is that the advantage Northern manufacturers have in terms of lower services intermediates cost is no longer enough to sustain a large number of manufacturers in the North at low levels of services trade cost. This is an interesting result since the loss of manufacturing or deindustrialisation in the North is not triggered by greater goods trade liberalisation (ϕ_M is kept constant in this set of results). Services

[60] In equilibrium, there will still be some manufacturing firms in the North. The presence of a large number of services firms means that there will still be demand for manufacturing intermediates.

trade liberalisation alone can trigger the deindustrialisation process.

The evolution of equilibrium wages is given in Figure 12. Trade costs are falling from the left to the right of the diagram. The boundary line at around $\phi_S = 0.2$, representing the services wage-break point, is the point where all skilled workers become fully employed in services. To the right of this point, skilled workers begin to earn a premium $w_K > w_L$. The biggest difference between w_K and w_L occurs at intermediate levels of τ_S (around $\phi_S = 0.43$ or $\tau_S = 1.23$).

Figure 12: Equilibrium Wages (Second Wave) North



At around $\phi_S = 0.8$ when the North suffers from a precipitous loss of manufacturing, K type workers in the North experience a wage spike. Consider what happens at this point. North's manufacturing sector has become so small that w_L becomes 1. In other words, some L type workers are reduced to working in agriculture, earning the same wage as O type workers. Much of manufacturing intermediates are now mainly imported from the South and they are subjected to trade cost. Due to the sudden relocation of manufacturing to the South, the CES price aggregate for manufacturing increases sharply in the North. Firms in the services sector demand three types of inputs - K-type workers, services intermediates and manufacturing intermediates. As the cost of manufacturing intermediates increases, firms substitute away from the last source of input towards the previous two. The effect of this is that w_K will rise sharply to restore equilibrium in the factor market.

4.5 Factor Responses and Changes in Endowments

International trade can be driven by inherent differences between countries giving rise to comparative advantage or by increasing returns to scale. Yet surprisingly little is known of the relative importance between the two except for the study by Davis and David E. Weinstein (1999). In this model, relative factor abundance (skilled, semi-skilled, unskilled workers) may determine where industries locate, but not necessarily so. If a region is relatively abundant in skilled labour, one would naturally expect that region to have a comparative advantage in the production of services. However, the overall cost of production depends on intermediate inputs as well as labour. The cost of intermediates in turn depends on the number of firms located in the region and the level of trade freedom. Comparative advantage therefore is not only due to the relative factor abundance but also the accidents of history - which industry agglomerates where first.

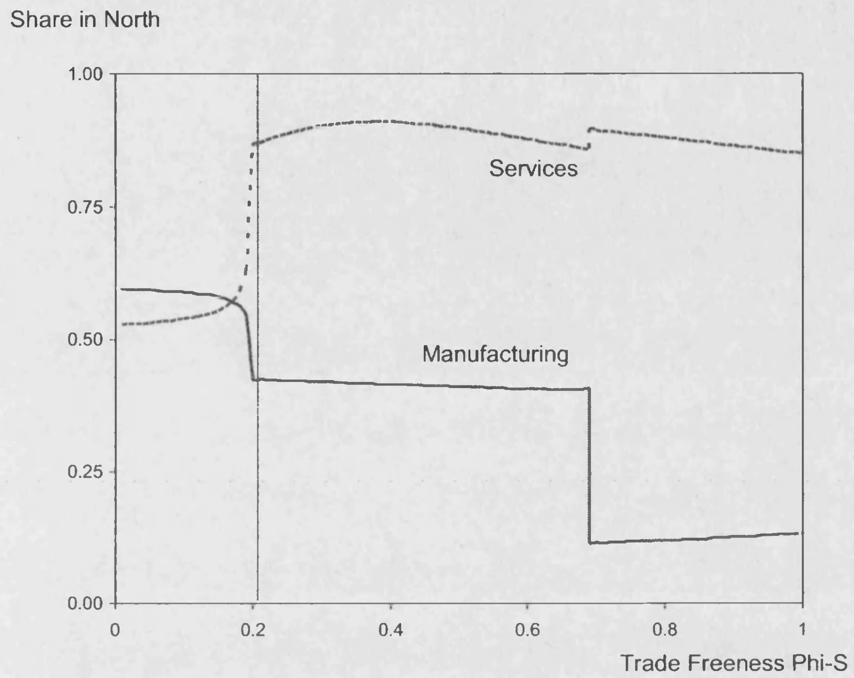
Davis and Weinstein (1999) find evidence that Economic Geography (or increasing returns as a motivation for trade) operates at a subnational scale but not for international trade which is still largely dominated by comparative advantage. However, as agglomeration of an industry often gives rise to economic rent, it is also possible that the composition of local factors begin to change in response to the agglomeration rent. Therefore, to the extent that factors shift endogenously in response to the kind of industries that agglomerate, there may in fact be more Economic Geography at work than evidence suggests. In a limited way, the next subsection discusses the endogeneity of Economic Geography and endowment-driven comparative advantage.

4.5.1 Comparisons with Neoclassical Theories

Consider the following thought experiment. K-type workers in the North earn a premium over L-types if services trade is sufficiently liberalised. Suppose there is an increase in K-type workers relative to L-types, what will happen to industrial locations and wages[61]? In this set of numerical solutions, the number of K-type workers in the North is increased by 10 units (or 10 per cent increase from baseline) and the number of L type workers is correspondingly decreased by 10 units (10 per cent decrease). For the world as a whole, this implies an increase in the number of K-type workers (by 5 per cent from baseline) and decrease in the number of L-types (by 5 per cent). The resulting industry shares are presented in Figure 13, which one can compare to Figure 11 showing the shares before the changes in endowment occur.

[61] One can think of the increase in K-types workers to arise from a process of skill acquisition, where the L-type workers acquire skills in response to the higher wages (or agglomeration rent) in the services sector. This paper does not model the dynamics of skills acquisition. Instead, the initial endowment is changed to reflect the acquisition of skills and a comparative static analysis is carried out.

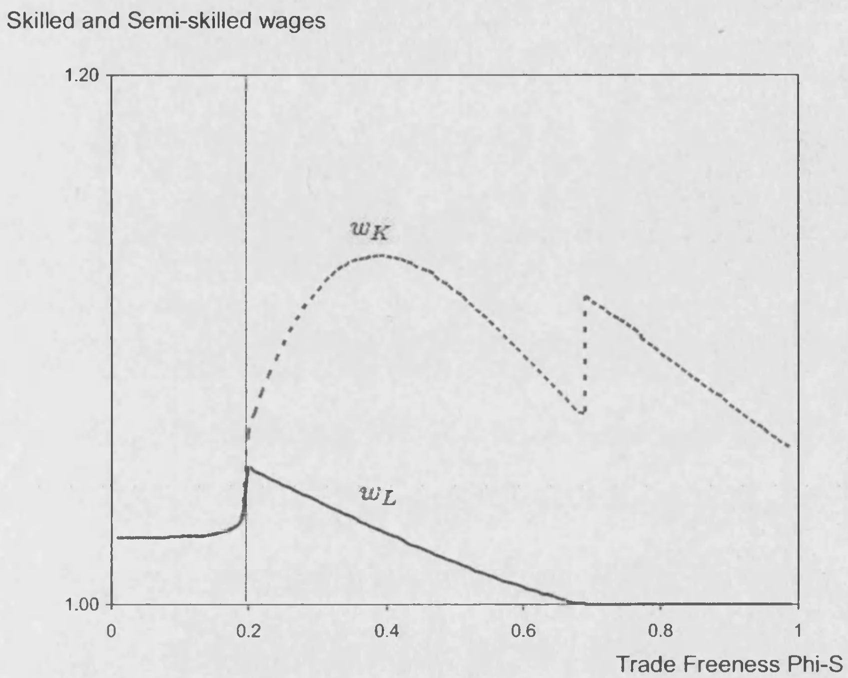
Figure 13: Tomahawk Diagram (Second Wave) with K Type Increase in North



As more semi-skilled workers become skilled, this further expands the services sector in the North by facilitating more services agglomeration. Compared to Figure 11, the comparative static shows that the share of North's manufacturing is even lower now at every level of services trade cost. Together, these present a quasi-Rybczynski effect, that is the expansion of the sector that uses the increased factor and a contraction of the other sector. Since inter-industry linkages are stronger than intra-industry ones, the cost of intermediates for manufacturing rises while that for services falls.

In equilibrium, it is possible that w_K rises even more, even though there

Figure 14: Equilibrium North Wages with Increase in K



This can be seen by comparing the wages shown in Figure 14 to that of Figure 12, without the increase in K-types. This is different from the prediction of neoclassical theories where one would expect the increase in a factor relative to another to either lead to decrease in relative wage (for a large country) or to have no effect at all (for a small country).

To draw the link between globalisation and developed economies' blue-collar wage decrease under the neoclassical model, economists have to demonstrate the price effect - that is, how trade liberalisation has reduced the relative price of goods which use blue-collar workers intensively [Feenstra and Hanson (1999)]. Simply put, the terms of trade have to move against a sector in order to account for the fall in the relative return of the factor used intensively in that sector in any neoclassical setting. This is simply a restatement of the

Stolpher-Samuelson theorem.

With increasing returns, input-output linkages and Economic Geography, this paper shows that a non-standard, or even surprising result, can emerge. The increase in the number of K-type workers in the North has led to a relative decrease in the (aggregated) price of services, yet this is accompanied by an increase in the returns to workers employed in that sector. In other words, the further agglomeration of services in the North results in the worsening of terms of trade or price effects, but at the same time leads to the relative increase in the factor return to skilled workers used in that sector. This is in sharp contrast to the Stolper-Samuelson prediction. The wage ratio, or relative wage, between skilled and unskilled becomes higher.

4.5.2 Offshoring and Wages

More recently, Gene Grossman and Esteban Rossi-Hansberg (2006), after further slicing the production of goods into a continuum of tasks, also show that offshoring of low-skilled work does not depress low-skilled wages if there are no terms of trade effects (such as in the case of a small country). In fact, offshoring of low-skilled work could increase low-skilled wages due to the productivity effect (or if the productivity effect is stronger than the terms of trade effect). However, the authors' framework is still very much neoclassical in nature.

In this paper with increasing returns to scale and input-output linkages, the prediction is again different. As more and more manufacturing firms migrate to the South, the sector becomes effectively offshored and intermediates have to be imported from the South. This represents a loss of intra-industry forward linkages, leading to an absolute decline in w_L in order to restore the equilibrium [see Figures 10 and 12]. By considering the effects forward linkages have on equilibrium wages, the conclusion about offshoring becomes less favourable for the less-skilled workers.

4.5.3 Worsening Inequality in the North

The upshot of all this is that the increase in skilled workers endowment in the North may worsen rather than ameliorate income inequality there. Indeed, Bound and Johnson (1995) point out that "One of the major puzzles about the wage structure during the 1980s [for the US] is why the returns to observed skills (education and experience) rose while the labour force has become more educated and older." There is indeed much evidence suggesting that skill-biased technology change can to a large extent explain this puzzle. However, skill-biased technological change does not account for one of the key stylised facts of the recent wave of globalisation, which is the deindustrialisation of many OECD countries.

Furthermore, Feenstra (1998) notes that "we should not assess the proximate cause of the decline in employment and wage of unskilled workers by attributing all within-industry shifts in labour to technology, and allowing trade to operate only via between industry shifts . . . as soon as trade in intermediates is permitted, as with outsourcing, then the changes in demand for labour within industry can occur due to trade as well. In fact, the whole distinction between 'trade' and 'technology' becomes suspect when we think of corporations shifting activities overseas."

4.5.4 Wage Inequality in the South

In the various sets of numerical solutions shown thus far, wage inequality appears only for the North. However, with a suitable choice of parameters, it is possible to show that inequality also arises in the South. The concentration of manufacturing in the South in the second wave of globalisation can lead to skilled and semi-skilled workers earning a wage premium above their unskilled counterparts (exactly the same process that occurred in the North during the first wave)[63]. Figure ?? is the South's counterpart to Figure 12. As trade costs fall from the left to the right of the diagram, it shows the effect of South's skilled and semi-skilled wages when manufacturing agglomerates in the second wave[64]. At around $\phi_S = 0.8$ (when manufacturing shifts dramatically to the South), there is an increase in the skilled and semi-skilled wages in the South over their unskilled counterparts. Again, note that this is entirely triggered by services trade liberalisation, which displaces manufacturing from the North. What began as sharp inequality between nations can end up becoming sharp inequality within nations.

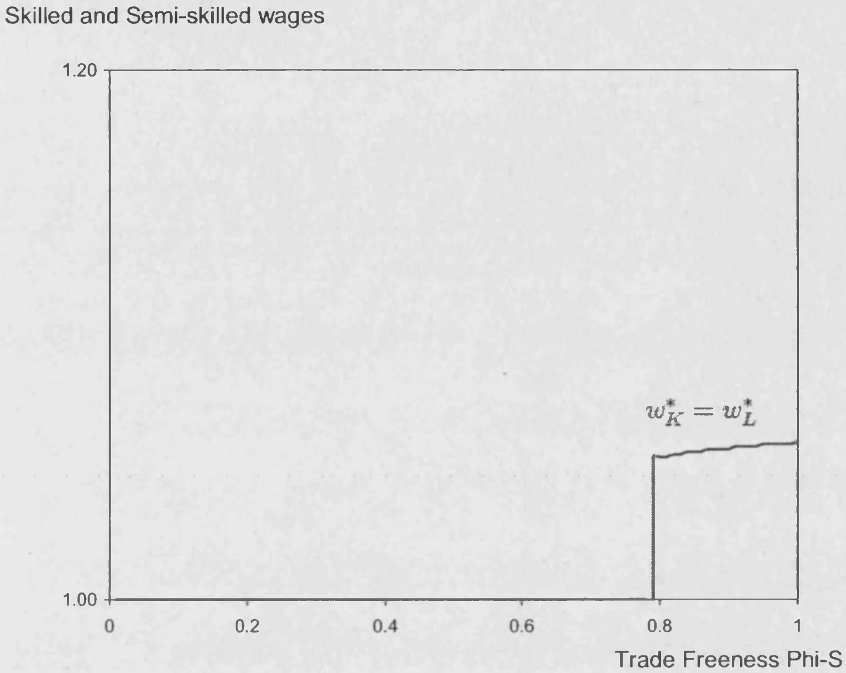
4.6 Conclusion

NEG models often rely on some simplifying assumptions, for example the presence of a homogeneous and costlessly traded good that equalises wages everywhere. In addition, due to the feedback mechanism in many NEG models, full analytical solutions are sometimes not possible. In its place, numerical

[63] Some economists have noted that inequality in developing countries can arise since the production that is offshored from developed countries allows the skilled workers in the developing countries to earn a wage premium above the rest of the population.

[64] An example of such a situation would be the income disparity between the urbanised manufacturing workers and the rural farmers in China.

Figure 15: Skilled and Semi-Skilled Wages in South (Second Wave)



solutions are often used to derive the economic understanding. In this paper, the tractability problem faced by a standard NEG model is made even more complicated with labour market segmentation, adjustments and wage dynamics. Furthermore, agglomeration forces operate in two different channels (inter and intra-industry) asymmetrically for given values of trade costs. Because of these complexities, only numerical solutions are provided. Naturally, numerical solutions have to be interpreted with a degree of caution. The results often change with parameter specifications. Without a full analytical solution, it is sometimes difficult to tell if the results are general enough for them to be plausible explanations of reality, a point this paper concedes. The advantage of numerical solution however is that it shows the rich equilibrium paths of locations and wages that can emerge in the model, and allows one to understand the complex dynamics of the globalisation process.

The contribution of this paper is that it reconciles several key stylised facts within a single Economic Geography framework. The story runs as follows. The improvement of communication technology in the recent wave of globalisation brings about a change in the specialisation of developed economies by allowing greater services agglomeration from their position of initial advantage in the first wave. Firstly, it results in a loss of workers from North's manufacturing sector due to competition for skilled workers. Secondly, greater services trade liberalisation results in the loss of forward linkages for North's manufacturing firms by allowing South's manufacturing sector to access services inputs cheaply. If services trade is free enough (such as in $\phi_S = 0.8$ in Figure 11), this can trigger a precipitous shift of manufacturing to the South or deindustrialisation (which implies greater offshoring). The loss of employment and forward linkages in North's manufacturing sector then reduces the wage of semi-skilled workers in the North. The increase in skilled labour in the North further accentuates this process and results in even greater inequality. By explicitly modelling the agglomeration process with labour markets constraints, and treating the globalisation process as two distinct waves, this paper shows how the shifts in economic geography can explain many stylised facts.This paper therefore provides a stylised understanding of the history of North-South industrial development and the patterns of trade.

4.7 Appendix C

4.7.1 Patterns of Specialisation in Major Developed and Developing Economies

Table 10: Current Account Balances (US Millions) of Goods (Top) and Services (Bottom) of Major OECD and Emerging Economies

Country	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Australia	-4223	-637	1766	-5355	-9751	-4715	1738	-5423	-15147	-18031	-13319	-9596
	-917	224	151	-852	128	938	736	1220	1818	511	533	837
Brazil	-3157	-5452	-6654	-6604	-1261	-698	2651	13121	24794	33666	44757	46115
	-7495	-8057	-9309	-9041	-6977	-7162	-7759	-4957	-4931	-4677	-8148	-9656
China	18050	19535	46222	46614	35982	34474	34017	44167	44652	58982	134189	
	-6093	-1984	-3398	-2777	-5341	-5601	-5933	-6784	-8573	-9699	-9391	
France	11004	14957	26955	24873	17627	-3300	3535	7506	3237	-4759	-28523	-37692
	14339	15107	16713	17347	18607	19805	17841	17128	15819	14554	13229	10382
Germany	65320	70890	71157	76047	69662	55986	88364	126607	145928	187445	189691	199775
	-53440	-51714	-48149	-51606	-57917	-55018	-54142	-43192	-50712	-51164	-51989	-47899
Italy	38710	53985	39995	36335	23488	9555	15587	13241	11209	11001	666	-11964
	6338	7183	7785	4891	1199	1076	16	-2868	-2668	1465	-649	-1851
India	-10721	-14635	-14787	-15613	-13705	-16496	-12047	-9556	-14633	-28075	-46915	-61239
	175	347	1276	2086	2238	3412	2865	4435	6401	13050	22225	29043
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Japan	131231	83595	101740	122116	123063	116506	70191	93675	105863	132903	94978	81200
	-57298	-62297	-54072	-49311	-54006	-47608	-43750	-42029	-35501	-38991	-27905	-20131
Korea	-4444	-14965	-3179	41627	28371	16954	13488	14777	21952	37569	32683	29214
	-2979	-6179	-3200	1024	-651	-2848	-3872	-8198	-7424	-8046	-13658	-18763
Netherlands	23790	22702	20925	20274	15916	17870	19174	18407	36528	41846	46328	48107
	1120	1970	3258	2511	2570	-2109	-2467	-1042	-699	4258	6796	5608
Russia	19816	21593	14913	16429	36014	60172	48121	46335	59860	85825	118364	139234
	-9637	-5384	-5945	-4083	-4284	-6665	-9131	-9886	-10894	-12693	-13894	-13812
Spain	-18612	-16048	-14219	-21459	-31938	-37076	-34569	-34409	-45008	-66670	-85261	-100593
	17420	18965	18239	19744	20498	19421	20559	21100	26212	26943	27677	27657
United Kingdom	-18973	-21413	-20216	-36140	-47001	-49920	-59358	-71570	-79387	-111575	-125044	-142739
	14135	17416	23130	24292	22052	20804	20804	24863	31259	47525	44828	53351
United States	-174170	-191000	-198428	-248221	-347819	-454690	-429519	-484955	-550892	-669578	-787149	-838271
	77786	86935	90155	82081	82729	74855	64393	61230	53977	57488	72778	79749

Countries with persistent goods deficits and services surpluses are: India, Spain, UK and United States. Countries with persistent goods surpluses and services deficits are: China, Germany, Indonesia, Japan, Korea and Russia. Countries with persistent goods surpluses: Netherlands. Countries with persistent services surpluses: France.

4.7.2 Model Parameters

Intra and Inter-industry Linkages The 1998 UK Input-Output Table with a total of 40 sectors covering both goods and services is used as a reference for parameter choice. The first 6 sectors spanning mostly primary products - agriculture, mining, food and tobacco, textile, wood, paper - are dropped. The utility sectors like construction and power generation are also dropped since they are infrastructural in nature (sectors 25-26). Services sectors that are mostly domestically oriented like hotels and restaurants (sector 28), real estate (sector 32), public administration and security (sector 37) are also dropped.

Of the remaining, sectors 7 to 24, spanning all industrial manufactured goods are taken as a whole to be the goods cluster. Sectors 27 to 40, spanning a whole range of services including wholesale and retail, finance, insurance, research and development, post & telecommunication, transport & storage and business services are considered as a whole to be the services cluster.

Ignoring the effects of production taxes, this paper checks for the intra- and inter-cluster linkages, taking the two broad clusters as defined. Checking the within and between cluster demands for intermediates, the goods cluster's inter and intra-industry linkages are 0.40 and 0.21 respectively. For the services cluster, the corresponding numbers are 0.43 and 0.06.

Instead of using cluster aggregate demands for intermediates, another measure would be to use the average intermediate demands of all defined sectors within the specified cluster. Taking the average for sectors in the goods cluster, the inter-industry and intra-industry linkages then become 0.44 and 0.24 respectively. Taking the average for sectors in the services cluster, the corresponding numbers are 0.50 and 0.07. The evidence therefore points to stronger inter-industry input-output linkages.

Elasticity David Hummels (1999) estimates that the elasticities of most goods are in the range of 3 to 8, with an average of 5.6. This implies an average mark-up of 22 per cent. This paper rounds down the elasticity to 5 for both services and manufacturing sectors, implying a mark-up of 25 per cent.

Consumption Intensities The consumption intensities for manufacturing and services are given as 0.4 and 0.2 respectively. The choice of parameters here have no bearing on the qualitative results except to note that they must be large enough (relative to endowments) so that one location cannot hold all firms. This is to ensure that the endowment constraints are binding enough to generate wage increases in equilibrium.

Summary The numerical solution is carried out using MATLAB. The baseline parameters are provided in the table below.

Table 11: Baseline Parameters for Numerical Solutions

Endowments	$K_N = L_N = U_N$	100
	$K_S = L_S = U_S$	100
Industry Linkages	α_M	0.40
	β_M	0.10
	α_S	0.40
	β_S	0.10
Elasticities	σ	5
	μ_M	0.40
	μ_S	0.20
Adjustment / Increment	ε_F	0.01
	ε_W	0.0001

5 Final Conclusion

While there is some broad agreement amongst mainstream economists of the benefits of globalisation, there is also a growing consensus that it is perhaps less useful to paint the process of globalisation in broad brush strokes. Globalisation is a complicated and multifaceted process. At the very least, it does create winners and losers, within and between nations. Judged by the strength of the anti-globalisation movement, the process of globalisation must have created real anxieties and increased insecurities, if not resulting in outright displacement. A key objective of research must therefore be to uncover the positive and normative aspects of globalisation and to address some of these concerns where possible.

Globalisation has not one, but many strands of narrative and research. While pointing to the gains of globalisation, new research suggests that globalisation works through many channels. The "new breed of models paint globalisation with a much finer brush."[65] - the effects on sectors, on firms, within firms and on different segments of society. With this over-arching view in mind, this thesis has highlighted three aspects of globalisation with three models.

The first model points to a key benefit of greater trade integration. In a monopolistically competitive, differentiated industry characterised by firm-heterogeneity and business cycle shocks, trade liberalisation can result in greater stability for price and output by allowing the productivity cutoffs to be equalised across economies. The greater stability then translates into di-

[65] The Economist, "The Great Unbundling", January 2007.

rect welfare gains for the risk-averse consumers. More pertinently, the result is not sensitive to the distribution of income. Greater price-output stability benefits all consumers and no one becomes worse off as a result of greater trade liberalisation. The model therefore presents a strong pro-globalisation argument.

The second model presents an Economic Geography environment with mobile capital, immobile workers, and productively heterogeneous firms. The results show that a small improvement in a location's productivity distribution (in terms of a rightward shift of the distribution) can result in high or even full concentration of industries there if trade is free enough. The other location becomes hollowed out (left only with agriculture) as all capital migrates to the more productive location. This may explain why some developing countries are highly ambivalent to the globalisation process. However, that model also provides some comfort to developing countries. Greater trade liberalisation can attract low sunk-cost industries and capital to locate there even if their productivity distribution is characterised by higher risk. This can help kick-start the industrialisation process. The level of trade costs may also explain the large discrepancies in the amount of FDI emerging economies receive.

Finally, the third model captures the key stylised facts that have emerged in the recent wave of globalisation - namely the deindustrialisation of some developed economies (offshoring), their transition into services specialisation, and within nation wage inequality in both developed and developing economies. By explicitly modelling the evolution of industrial locations and the agglomeration process driven by input-output linkages, the model provides predictions different from neoclassical theories. In particular, the model shows how globalisation can lead to the loss of manufacturing employment, stagnation or even decline of blue-collar wages in the developed countries.

In other words, globalisation is not always pareto improving. The fact that globalisation creates winners and losers, and the backlash it currently generates, underscores this point. Is the process of globalisation therefore as inevitable as the laws of gravity as Kofi Anan suggests? The era before the First World War was the last high point in globalisation. Feenstra (1998) in fact notes that if one uses merchandise trade as a percentage of GDP to be the proxy for globalisation, many industrialised countries today are only as globalised as on the eve of the First World War. What happened after the war - the Great Depression, the rise of competitive protectionism and economic nationalism - are well documented. The lessons of history inform us that globalisation is not necessarily an inevitable or irrevocable process. As Kofi Anan also notes, "Globalisation is a fact of life, but I believe we have underestimated its fragility." How the process of globalisation should be managed to provide the greatest benefits for the greatest number of people, remains a keenly debated question.

References

[1] Alessandria, George and Choi, Horag (2007), "Do Sunk Costs of Exporting Matter for Net Export Dynamics", Quarterly Journal of Economics.

[2] Alfaro, Laura; Kalemli-Ozcan, Sebnem; Volosovych, Vadym (2005), "Why Doesn't Capital Flow from Rich to Poor Countries? An Empirical Investigation", NBER Working Paper No.11901.

[3] Bairoch, Paul.(1982), "International Industrialisation Levels from 1750 - 1980", Journal of European History 2, pp. 268 - 333.

[4] Bairoch, Paul.(1989), "European Trade Policy, 1815 - 1914", The Cambridge Economic History of Europe (Vol.8), Cambridge University Press.

[5] Balwin, Richard E.; Forslid, Richard; Martin, Phillippe; Ottaviano, Gianmarco I.P.; Robert-Nicoud, Frederic., "Economic Geography & Public Policy", Princeton University Press.

[6] Baldwin, Richard E. and Martin, Philippe (1999), "Two Waves of Globalisation: Superficial Similarities, Fundamental Differences", NBER Working Paper No. 6904.

[7] Baldwin, Richard E.; Martin, Philippe and Ottaviano, Gianmarco I.P. (2001), "Global Income Divergence, Trade and Income Divergence: The Economic Geography of Growth Takeoffs", Journal of Economic Growth, Vol 6, pp. 5 - 37.

[8] Baldwin, Richard E. and Okubo, Toshihiro (2005), "Heterogeneous Firms, Agglomeration and Economic Geography: Spatial Selection and Sorting", Journal of Economic Geography 6, pp. 323-346.

[9] Baldwin, Richard E. and Okubo, Toshihiro (2006), "Agglomeration, Offshoring and Heterogeneous Firms", CEPR Discussion Paper DP5663.

[10] Baldwin, Richard E. and Robert-Nicoud (2006), "Trade and Growth with Heterogeneous Firms", NBER Working Paper No. 12326.

[11] Banerjee, Abhijit V. and Duflo, Esther (2004), "Growth Theory through the Lens of Development Economics", Handbook of Development Economics, Vol. 1a, pp. 473-552

[12] Batra, Raveendra N. and Russell, William R. (1974), "Gains from Trade and Uncertainty", The American Economic Review, Vol.64, No.6, pp. 1040-1048.

[13] Behrens, Kristian and Murata, Yasusada (2005), "A Functional Separability Approach to Monopolistic Competition: The Dixit-Stiglitz Model Reconsidered", Discussion Paper No. 807, Interfaces for Advanced Economic Analysis, Kyoto University.

[14] Behrens, Kristian and Murata, Yasusada (2007), "General Equilibrium Models of Monopolistic Competition: A New Approach", Journal of Economic Theory (forthcoming).

[15] Bejan, Maria (2006), "Trade Openness and Output Volatility", unpublished.

[16] Bernard, Andrew B.; Eaton, Jonathan; Jensen, J. Bradford and Kortum, Samuel (2003), "Plants and Productivity in International Trade".

[17] Bernard, Andrew B.; Redding, J. Stephen and Schott, Peter K. (2007), "Comparative Advantage and Heterogeneous Firms", Review of Economic Studies Vol 74, pp.31 - 66.

[18] Bernard, Andrew B.; Redding, J. Stephen and Schott, Peter K. (2006), "Multi-Product Firms and Trade Liberalization".

[19] Bound, John and Johnson, George (1992), "Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations", American Economic Review, Vol.82 No.3, pp. 371 - 392.

[20] Bound, John and Johnson, George (1995), "What are the Causes of Rising Wage Inequality in the United States", Federal Bank of New York Economic Policy Review.

[21] Brenner, Neil (1998), "Global Cities, Glocal States: Gobal City Formation and State Territorial Restructuring in Contemporary Europe", Review of International Political Economy 5:1, pp 1-37.

[22] Caselli, Francesco and Feyrer, James (2007), "The Marginal Product of Capital", London School of Economics, Quarterly Journal of Economics, May 2007, pp. 535 - 568.

[23] Chaney, Thomas (2006), "Distorted Gravity: Heterogeneous Firms, Market Structure and Geography of International Trade", MIT, unpublished.

[24] Davis, Donald R. (1998), "The Home Market, Trade, and Industrial Structure", The American Economic Review Vol 88, pp 1624-1276.

[25] Davis, Donald R. and Weinstein, David E. (1999), "Economic Geography and Regional Production Structure: An Empirical Investigation", European Economic Review Vol 43, pp 379 - 407.

[26] Demidova, Svetlana (2006), "Productivity Improvements and Falling Trade Cost: Boon or Bane?", International Economic Review, 2nd Revision.

[27] Dennis, Robert (1978), "The Decline of Manufacturing Employment in Greater London: 1966-1974", Urban Studies 15:1, pp. 63 - 73.

[28] Dixit, Avinash (1987), "Trade and Insurance with Moral Hazard", Journal of International Economics 23, pp. 201-220.

[29] Dixit, Avinash (1989), "Trade and Insurance with Imperfectly Observed Outcomes", The Quarterly Journal of Economics, Vol 104, No.1, pp. 195-203.

[30] Dixit, Avinash K. and Stiglitz, Joseph E. (1977), "Monopolistic Competition and Optimum Product Diversity", The American Economic Review, 67(3), pp 297 - 308.

[31] Falvey, Rod; Greenaway, David and Yu, Zhihong (2006), "Extending the Melitz Model to Asymmetric Countries", GEP, University of Nottingham.

[32] Ghironi, Fabio and Melitz, Marc J. (2004), "International Trade and Macroeconomic Dynamics with Heterogeneous Firms", CEPR Discussion Paper No. 4595.

[33] Griffith-Jones, Stephany and Leape, Jonathan (2002), "Capital Flows to Developing Countries: Does the Emperor Have Clothes?", QEH Working Paper Series, Working Paper Number 89.

[34] Gripaios, Peter (1977), "Industrial Decline in London: An Examination of Its Causes", Urban Studies 14:2, pp. 181 - 189.

[35] Grossman, Gene and Rossi-Hansberg, Esteban, "Trading Task: A Simple Theory of Offshoring", Princeton University 2006.

[36] Hummels, David (1999), "Towards a Geography of Trade Cost", Purdue University.

[37] Feenstra, Robert (1998), "Integration of Trade and Disintegration of Production in the Global Economy", The Journal of Economic Perspective, Vol. 12, No.4, pp. 31 - 50..

[38] Feenstra, Robert C. and Hanson, Gordon H. (1999), "The Impact of Outsourcing and High-Technology Capital on Wages: Estimates for the United States, 1979 - 1990", The Quarterly Journal of Economics, Vol. 114 No. 3, pp. 907 - 940.

[39] Fujita, Masahisa; Krugman, Paul and Venables, Anthony J., "The Spatial Economy: Cities, Regions and International Trade"

[40] Fujita, Masahisa; Krugman, Paul and Mori, Tomoya (1999), "On the Evolution of Hierarchical Urban Systems", European Economic Review 43, pp. 209 - 251.

[41] Harrigan, James. and Venables, Anthony J. (2004), "Timeliness, Trade and Agglomeration", Centre of Economic Performance (LSE).

[42] Helpman, Elhanan; Melitz, Marc J. and Yeaple. Stephen R. (2004), "Export versus FDI with Heterogeneous Firms", The American Economic Review, Vol 94 No.1, pp. 300 - 317.

[43] Helpman, Elhanan and Razin, Assaf (1978), "Uncertainty and International Trade in the Presence of Stock Markets", The Review of Economic Studies, Vol.45, No.2, pp. 239-250.

[44] Hirsh, Seev and Lev, Barauch (1971), "Sales Stabilization Through Export Diversification", The Review of Economics and Statistics, Vol.53, No.3, pp.270-277.

[45] Krugman, Paul R.(1979), "Increasing Returns, Monopolistic Competition, and International Trade", Journal of International Trade, 9(4), pp. 469 – 479.

[46] Krugman, Paul R.(1980), “Scale Economies, Product Differentiation, and Pattern of Trade”, The American Economic Review 70, pp. 950 – 959.

[47] Krugman, Paul R.(1991), “Increasing Returns and Economic Geography”, Journal of Political Economy 99, pp. 483 – 499.

[48] Krugman, Paul R. and Venables, Anthony J. (1995), “Globalization and the Inequality of Nations”, The Quarterly Journal of Economics, Vol. 110, No.4, pp 857-880.

[49] Kucera, David and Milberg, William (2002), “Trade and the Loss of Manufacturing Jobs in the OECD: New Factor Contents Calculations for 1978 - 1995”, International Institute for Labour Studies, Discussion paper 135.

[50] Martin, Philippe and Rogers, Carol Ann, “Industrial Location and Public Infrastructure”, Journal of International Economics 39, pp.355 - 351.

[51] Melitz, Marc J. (2003), “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity,” Econometrica 71, pp. 1695-1725.

[52] Melitz, Marc J. and Ottaviano, Gianmarco I.P (2005), “Market Size, Trade, and Productivity”, NBER Working Paper No.11393..

[53] Montagna, Catia (1995), “Monopolistic Competition with Firm-Specific Cost”, Oxford Economics Paper 47, pp. 318 - 328.

[54] Mundell, Robert (1957), "International Trade and Factor Mobility", American Economic Review, Vol 47 No.3, pp. 321 - 335.

[55] Neary, Peter J. (2001), “Of Hype and Hyperbolas: Introducing The New Economic Geography”, Journal of Economic Literature, Vol 39 No.2, pp. 536 - 561.

[56] Newbery, David M.G., and Stiglitz, Joseph E. (1984), "Pareto Inferior Trade", The Review of Economic Studies 51, pp. 1 – 12.

[57] Obstfeld, Maurice and Rogoff, Kennedy (2000), "The Six Major Puzzle in International Economics: Is There a Common Cause?", NBER Economic Annual 2000.

[58] Redding, Stephen J. and Venables, Anthony J. (2003), "Economic Geography and International Inequality", Journal of International Economics Vol 62, pp 53-82.

[59] Reinhart, Carmen and Rogoff, Kenneth S. (2004), "Serial Default and the "Paradox" of Rich to Poor Capital Flow", NBER Working Paper No. 10296.

[60] Robert-Nicoud (2006), "Agglomeration and Trade with Input-Output Linkages and Capital Mobility", Spatial Economic Analysis (forthcoming).

[61] Rodrik, Dani (1998), "Why do More Open Economies have Bigger Governments", The Journal of Political Economy, Vol. 106, No.5, pp 997-1032.

[62] Romo, Frank P. and Schwartz, Michael (1995), "The Structural Embeddedness of Business Decisions: The Migration of Manufacturing Plants in New York State, 1960 to 1985", American Sociology Review, Vol 60, No 6, pp. 874 - 907.

[63] Sassen, Saskia (1991), "The Global City, New York, London, Tokyo", Princeton University Press.

[64] Tornell, Aaron and Velasco, Andres (1992), "The Tragedy of the Commons and Economic Growth: Why Does Capital Flow from Poor to Rich Countries", The Journal of Political Economy, Vol 100, No.6, pp.1208-1231.

[65] Tybout, James R. (2000), "Manufacturing Firms in Developing Countries: How Well Do They Do, and Why?", Journal of Economic Literature, Vol. 38, No.1, pp 11-44.

[66] Venables, Anthony J. (1996), "Equilibrium Locations of Vertically Linked Industries", International Economic Review, Vol. 37, No.2, pp 341 - 359.

[67] Venables, Anthony J. (2006), "Shifts in Economic Geography and their Causes", CEP Discussion Paper No. 0767.