

The Monopolistic Competition Revolution in Retrospect

Steven Brakman and Ben J. Heijdra, Editors



CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge CB2 1RP, United Kingdom

CAMBRIDGE UNIVERSITY PRESS

The Edinburgh Building, Cambridge, CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

<http://www.cambridge.org>

© Cambridge University Press 2004

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2004

Printed in the United Kingdom at the University Press, Cambridge

Typeface Plantin 10/12 pt. System L^AT_EX 2_ε [TB]

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication data

The monopolistic competition revolution in retrospect/Steven Brakman
and Ben J. Heijdra, editors.

p. cm.

Includes bibliographical references and index.

ISBN 0-521-81991-1

1. Competition, Imperfect. 2. Competition. 3. Monopolies.

I. Brakman, Steven. II. Heijdra, Ben J.

HB238.M66 2003

338.6'048-dc21

2003048569

ISBN 0 521 81991 1 hardback

The publisher has used its best endeavours to ensure that the URLs for external websites referred to in this book are correct and active at the time of going to press. However, the publisher has no responsibility for the websites and can make no guarantee that a site will remain live or that the content is or will remain appropriate.

Contents

<i>List of contributors</i>	<i>page xi</i>
<i>Preface</i>	<i>xiii</i>
1 Introduction	1
STEVEN BRAKMAN AND BEN J. HEIJDR	
1.1 Introduction	1
1.2 Precursory thoughts on imperfect competition	3
1.3 Monopolistic competition in the 1930s	7
1.4 The second monopolistic revolution	12
1.5 Structure of the book	27
 Part I Underground classics	 47
2 Monopolistic competition and the capital market	49
JOSEPH E. STIGLITZ	
2.1 Introduction	49
2.2 The model	54
2.3 The market solution	57
2.4 Competitive versus optimal size of risky industry	58
2.5 Correlated returns: the competitive analysis	60
2.6 Increasing marginal entrance costs	63
2.7 Reinterpretation in partial equilibrium terms	64
2.8 Concluding comments	67
 3 Monopolistic competition and optimum product diversity (May 1974)	 70
AVINASH K. DIXIT AND JOSEPH E. STIGLITZ	
3.1 Introduction	70
3.2 The basic model	71
3.3 Monopolistically competitive equilibrium	76
3.4 Constrained optimality	78
3.5 Unconstrained optimum	82
3.6 Possible generalisations	86
3.7 Concluding remarks	87

4	Monopolistic competition and optimum product diversity (February 1975)	89
	AVINASH K. DIXIT AND JOSEPH E. STIGLITZ	
	4.1 Introduction	89
	4.2 The demand for variety	92
	4.3 The constant elasticity case	97
	4.4 Diversity as a public good	103
	4.5 Variable elasticity utility functions	107
	4.6 Asymmetric cases	112
	4.7 Concluding remarks	119
	Part II Current perspectives	121
5	Some reflections on theories and applications of monopolistic competition	123
	AVINASH K. DIXIT	
	5.1 Introduction	123
	5.2 Alternative models of monopolistic competition	125
	5.3 Some themes from the conference papers	131
	5.4 Concluding remarks	132
6	Reflections on the state of the theory of monopolistic competition	134
	JOSEPH E. STIGLITZ	
	6.1 Introduction	134
	6.2 Chamberlin's theory of monopolistic competition	135
	6.3 Alternative modelling approaches	137
	6.4 Normative analysis	140
	6.5 Contestability doctrines	142
	6.6 Schumpeterian competition	144
	6.7 Conclusions	145
7	Dixit-Stiglitz, trade and growth	149
	WILFRED J. ETHIER	
	7.1 Introduction	149
	7.2 Trade	150
	7.3 Growth	152
	7.4 Concluding remarks	155
	Part III International trade	157
8	Monopolistic competition and international trade theory	159
	J. PETER NEARY	
	8.1 Introduction	159
	8.2 The Dixit-Stiglitz model and trade theory	160
	8.3 An extension	171
	8.4 Lacunae	174
	8.5 Conclusion	179

9	Monopolistically competitive provision of inputs: a geometric approach to the general equilibrium	185
	JOSEPH FRANCOIS AND DOUGLAS NELSON	
9.1	Introduction	185
9.2	National production externalities in autarky: model I	187
9.3	National production externalities with trade: model II	195
9.4	International production externalities: models III and IV	196
9.5	Factor market flexibility and industrialisation patterns	202
9.6	Globalisation and wages	203
9.7	Summary	206
	Part IV Economic geography	211
10	The core-periphery model: key features and effects	213
	RICHARD E. BALDWIN, RIKARD FORSLID, PHILIPPE MARTIN, GIANMARCO I.P. OTTAVIANO AND FREDERIC ROBERT-NICOUD	
10.1	Introduction	213
10.2	The standard core-periphery model	214
10.3	The long-run equilibria and local stability	218
10.4	Catastrophic agglomeration and locational hysteresis	221
10.5	The three forces: intuition for the break and sustain points	222
10.6	Caveats	230
10.7	Global stability and forward-looking expectations	231
10.8	Concluding remarks	234
11	Globalisation, wages and unemployment: a new economic geography perspective	236
	JOLANDA J. W. PEETERS AND HARRY GARRETSEN	
11.1	Introduction	236
11.2	The model	238
11.3	Short-run implications of globalisation	244
11.4	Long-run implications of globalisation	254
11.5	Conclusions	257
12	Empirical research in geographical economics	261
	STEVEN BRAKMAN, HARRY GARRETSEN, CHARLES VAN MARREWIJK AND MARC SCHRAMM	
12.1	Introduction	261
12.2	Empirical research in economic geography	263
12.3	Germany	271
12.4	Conclusions	281
13	The monopolistic competition model in urban economic geography	285
	J. VERNON HENDERSON	
13.1	Introduction	285
13.2	Krugman's application of Dixit-Stiglitz	286

13.3	Issues with the core-periphery model and its derivatives	289
13.4	Dixit–Stiglitz as micro-foundations for agglomeration	296
13.5	Recent developments	297
Part V Economic growth		305
14	Monopolistic competition and economic growth	307
SJAK SMULDERS AND THEO VAN DE KLUNDERT		
14.1	Introduction	307
14.2	The model	309
14.3	Growth through variety expansion	313
14.4	Growth through in-house R&D	319
14.5	Growth with variety expansion and in-house R&D	324
14.6	Conclusions	329
15	Convergence and the welfare gains of capital mobility in a dynamic Dixit–Stiglitz world	332
SJAK SMULDERS		
15.1	Introduction	332
15.2	A two-country endogenous growth model	335
15.3	Balanced trade versus capital mobility	341
15.4	How does monopolistic competition affect convergence?	345
15.5	Welfare	349
15.6	Conclusions	352
	Appendix	353
16	A vintage model of technology diffusion: the effects of returns to diversity and learning-by-using	356
HENRI L. F. DE GROOT, MARJAN W. HOFKES AND PETER MULDER		
16.1	Introduction	356
16.2	The model	358
16.3	Solution of the model	363
16.4	Comparative static characteristics	367
16.5	Conclusion	370
Part VI Macroeconomics		373
17	Monopolistic competition and macroeconomics: theory and quantitative implications	375
RUSSELL W. COOPER		
17.1	Motivation	375
17.2	A theory structure	375
17.3	Quantitative analysis: response to technology shocks	379
17.4	Policy implications	387
17.5	Conclusion	396

18	Does competition make firms enterprising or defensive?	399
	JAN BOONE	
	18.1 Introduction	399
	18.2 The model	402
	18.3 Partial equilibrium model: appropriability	406
	18.4 General equilibrium model: downsizing	408
	18.5 Concluding remarks	413
	Appendix	414
19	Rationalisation and specialisation in start-up investment	417
	CHRISTIAN KEUSCHNIGG	
	19.1 Introduction	417
	19.2 The model	420
	19.3 General equilibrium	426
	19.4 Social optimum	433
	19.5 Conclusions	438
	Appendix	439
20	Industrial policy in a small open economy	442
	LEON J. H. BETTENDORF AND BEN J. HEIJDR	
	20.1 Introduction	442
	20.2 The model	444
	20.3 Macroeconomic effects of industrial policy	454
	20.4 Welfare effects of industrial policy	467
	20.5 Conclusions	472
	Appendix	473
	<i>Index</i>	485

1 Introduction

Steven Brakman and Ben J. Heijdra

1.1 Introduction

In speaking of theories of monopolistic or imperfect competition as ‘revolutions,’ I know in advance that I shall provoke dissent. There are minds that by temperament will define away every proposed revolution. For them it is enough to point out that Keynes in 1936 had some partial anticipator in 1836. Newton is just a guy getting too much credit for the accretion of knowledge that covered centuries. A mountain is just a high hill; a hill merely a bulging plain. Such people remind me of the grammar-school teacher we all had, who would never give 100 to a paper on the ground that ‘No one is perfect.’ (Samuelson, 1967, p. 138)

Edward Hastings Chamberlin is the author of one of the most influential works of all time in economic theory – *The Theory of Monopolistic Competition*, which entered its eighth edition in 1962. Along with Lord Keynes’s *General Theory*, it wrought one of the two veritable revolutions in economic theory in this century. (Dust cover text of Kuenne, 1967)

Although we stress the importance of the contribution by Avinash Dixit and Joseph Stiglitz (1977) throughout this book, the history of monopolistic competition is much longer than the past twenty-five years or so and goes back at least seventy years. The success of the Dixit–Stiglitz model of monopolistic competition might have come as a surprise to students of the history of economic thought, as it was by no means the first attempt to deal with imperfect markets or monopolistic competition. However, where the earlier attempts failed the Dixit–Stiglitz approach turned out to be very successful and has the potential ‘for classic status’ (see Neary,¹ chapter 8 in this volume).

In this introduction we will briefly review the two waves of literature on monopolistic competition theory, namely the one that started in 1933 and the one that commenced in 1977. The claim of this book is that the second attempt to model monopolistic competition was far more successful than the first, essentially because the second attempt introduced

We thank Avinash Dixit for comments on an earlier draft.

¹ According to Peter Neary, ‘the first step on the road to classic status [is]: to be widely cited but never read. (The second step, to be widely quoted but never cited.)’

a formalisation that had all the relevant characteristics of monopolistic competition but was still relatively easy to handle.

This collection of papers will show that the re-formulation by Dixit and Stiglitz has contributed significantly to many areas of research; the main ones being international trade theory, macroeconomics, growth theory and economic geography. But even today the concept of monopolistic competition is not always appreciated. As David Kreps puts it in his influential micro textbook 'were it not for the presence of this theory in most lower level texts we would ignore it here altogether' (1990, p. 344). Kreps dismisses monopolistic competition as being too unrealistic, and challenges his readers to come up with at least one sector that could convincingly be described by monopolistic competition. This collection of essays, however, takes for granted that the Dixit–Stiglitz reformulation of monopolistic competition has become very successful, and asks why that is the case. This does not mean that the authors of the essays are uncritical about the model. The aim of this collection is to show why the model has become mainstream in such a short period of time and what we can expect from future developments regarding the modelling of imperfect markets.

This introductory chapter is organised as follows. In section 1.2 we briefly discuss the literature predating the first monopolistic competition revolution. This literature strongly hinted at the importance of increasing returns to scale and imperfect market forms but was unable to come up with a satisfactory model in which both phenomena could play a meaningful role.

In section 1.3 we briefly discuss (what we call) the first monopolistic competition revolution, namely the one that was started by Edward Hastings Chamberlin and Joan Robinson in the 1930s. We show that by the mid-1960s most (but not all) leading economists had come to the conclusion that the Chamberlin–Robinson revolution had essentially failed. In our view, there are two reasons for this lack of acceptance of the theory. First, the *timing* of the first revolution was unfortunate in that it coincided with the Great Depression and the emergence of the Keynesian revolution in macroeconomics. Second, and perhaps more importantly, Chamberlin and co-workers failed to come up with a canonical model embodying the key elements of the theory. It was not so much Chamberlin's ideas that were rejected but rather his *modelling approach* that was deemed to be unworkable.

In section 1.4 we turn to the second monopolistic competition revolution, namely the successful one that was started in the mid-1970s by Dixit, Stiglitz and Michael Spence. The timing of this second revolution was much better. The events in the world – the petroleum cartel, high

inflation, productivity slowdown, etc. – made the profession painfully aware of the limitations of the paradigm of perfect competition, and made it more receptive to theories that departed from that paradigm in all its dimensions, i.e. returns to scale, uncertainty and information, strategic behaviour, etc. In addition, the second revolution caught on because Dixit and Stiglitz managed to come up with a canonical model of monopolistic competition. We present a very simple version of the Dixit–Stiglitz model and show how it manages to capture the key Chamberlinian insights.

Finally, in section 1.5 we present a broad overview of the chapters in this book.

1.2 Precursory thoughts on imperfect competition²

By the end of the nineteenth century two market forms dominated the discussion of economic analysis, namely monopoly and perfect competition. The former assumes a single firm with exclusive control over its output and the market, resulting in profits that are larger than in any other market form. In contrast, the latter assumes a large number of sellers of a homogeneous product, where each individual firm has no control over its price. Free entry and exit of firms ensures that long-run profits are zero. Perfect competition was introduced to show that in some sense it is optimal and in fact represents an end-state, meaning that competition between buyers or sellers has come to an end and neither party can increase utility or profits by changing its behaviour. Changes occur only if exogenous variables change, but the question then becomes how fast and under what circumstances the new equilibrium will be reached. Competition might not actually lead to the blissful state but market forces are always pointing the economy in the right direction.³ Monopoly by contrast maximises profits of the firm but from a social point of view is sub-optimal.

This state of affairs is reflected in Alfred Marshall's *Principles of Economics*, that presented these two market forms as the basic analytical tools to analyse markets. Other market forms are hybrids in between these two

² Our historical overview is rather succinct owing to space considerations. Interested readers are referred to Triffin (1940), Eaton and Lipsey (1989, pp. 761–6) and Archibald (1987, pp. 531–4) for more extensive surveys.

³ As Arrow and Debreu showed, in general the conditions for a unique and stable (Walrasian) equilibrium are that (1) production is subject to constant or diminishing returns to scale, (2) commodities are substitutes (meaning that a price increase raises the demand for other products), (3) external effects are absent and (4) there is a complete forward market for all goods. Assumptions (1) and (3) in particular are dropped in monopolistic competition.

polar cases.⁴ Mainstream economics did not bother too much to analyse imperfect market forms, because 'the large majority of cases that occur in practice are nothing but mixtures and hybrids of these two' (Schumpeter, 1954, p. 975).

However, Marshall was aware that other market forms were not simple combinations of perfect competition and monopoly. The special nature of imperfect markets were conveyed to him in the form of the duopoly models developed by Cournot, Bertrand and Edgeworth in the second half of the nineteenth century. The analysis of Cournot (1838) was particularly important for him, as it handed him the apparatus to analyse market forms in the first place. The problem with these models was that the results depended very much on special assumptions. Although Marshall did not develop his own theory of imperfect competition, his awareness of the so-called 'Special Markets' paved the way for later theories of imperfect competition developed by Chamberlin and Robinson.

Notwithstanding some lip-service to the theory of imperfect competition, perfect competition dominated the analysis during this time and other market forms were considered to be 'imperfect'. However, in perfect competition, where each seller or buyer has no influence on market prices, there is no longer room for individual competition, and forces leading to *industry* growth are absent. The difficulty was then to reconcile the theory of the market and that of the individual firm. Simple observation of reality often contradicted the conclusions of (partial) supply and demand analysis: diminishing returns for the individual firm is not an obstacle to expand production. And average costs are diminishing at the point where firms stop expanding output. This state of affairs troubled Marshall, as decreasing (average) cost curves are incompatible with perfect competition. Marshall tried to solve this by introducing diminishing returns for the individual firm (for individual firms, production factors are in fixed supply), and external economies for the whole industry. The introduction of external economies of scale at the industry level ensured that the competitive equilibrium could be rescued. The central idea is that external economies of scale create an interdependence between supply curves; the combined supply of all firms reduces industry costs and ensures that the combination of lower prices and increased supply can be an equilibrium. External economies of scale are compatible with an industry equilibrium, because an increase in demand will still increase the price for *individual*

⁴ However, according to Schumpeter, Marshall 'had no theory of monopolistic competition. But he pointed toward it by considering a firm's Special Market' (Schumpeter, 1954, p. 840).

firms, as the marginal cost curve of each firm is upward sloping and each firm is operating at the minimum of its average cost curve. The price increase could stimulate new firms to enter the market, reducing (average) costs and raising combined supply. With internal economies of scale a market equilibrium is not possible as each individual firm can always undercut its rivals.

According to Marshall whether or not external economies could be encountered in practice depended on the general characteristics of an industry and the environment of the industry, like the localisation of an industry. In Marshall's words:

subsidiary trades grow up in the neighbourhood, supplying it with implements and materials, organizing its traffic, and in many ways conducing to the economy of its material . . . the economic use of expensive machinery can sometimes be attained in a very high degree in a district in which there is a large aggregate production of the same kind, . . . subsidiary industries devoting themselves each to one small branch of the process of production, and working it for a great many of their neighbours, are able to keep in constant use machinery of the most highly specialized character, and to make it pay its expenses. (Marshall, 1920, p. 225)

In modern jargon the linkages described in this quotation are so-called backward and forward linkages; the backward linkage is that firms use other firms' output as intermediate production factors, the forward linkage is that its own product is also used as an intermediate production factor by others.⁵

Furthermore, according to Marshall a thick labour market also benefits firms:

Employers are apt to resort to any place where they are likely to find a good choice of workers with the special skill which they require; while men seeking employment naturally go to places where there are many employers who need such skill as theirs and where therefore it is likely to find a good market. (Marshall, 1920, pp. 225–6)

These factors combined explain industry growth and show why:

the mysteries of the trade become no mysteries; but are as it were in the air . . . if one man starts a new idea, it is taken up by others and combined with suggestions of their own; and thus it becomes the source of further new ideas. (Marshall, 1920, p. 225)

⁵ The quote from Marshall merely seems to shift the problem to a different level, in the sense that external economies of scale in one industry must be explained by internal economies of scale in an upstream or downstream industry linked to it, and that raises doubts about sustainability of perfect competition in that other industry.

For Marshall, however, his analysis of external economies created an additional problem, because he thought that internal economies of scale were at least as important as external economies (Blaug, 1997). In the presence of internal economies of scale the growth of an industry would benefit the largest firms (and create monopolies) and thus change the competitive forces within such an industry. Marshall had to introduce the concept of the representative firm to deal with this incompatibility. By introducing the representative firm, perfect competition and (external) economies of scale could be made consistent. But again in this case, as with perfect competition, strategic interaction between firms has been assumed away because firms are by assumption 'representative' for the whole industry.

But the consistency problems in Marshall's analysis of the market were not solved even by the representative firm. Marshall's famous period analysis assumed that in the long run the supply curve was a straight line. And this means that *in the long run* the volume of production of an individual firm is indeterminate: there is no unique intersection of the supply curve and a given price. So, Marshall's theory of perfect competition has no way of dealing with situations where the (long-run) marginal costs are constant (or declining in the presence of economies of scale). This state of affairs was most poignantly put forward by Sraffa (1926). According to Sraffa market imperfections due to returns to scale are not simple frictions, 'but are themselves active forces which produce permanent and even cumulative effects'. And he added yet another problem. Declining marginal costs would imply that the market is served by a single firm. But, according to Sraffa, in practice firms operate under declining marginal costs without monopolising the whole market. According to him, the combination of a declining supply curve and a negatively sloped demand curve limits the size of production. The idea behind a declining demand curve is that buyers are not indifferent between different suppliers. Each firm has his own *special market*; products are usually imperfect substitutes and have their own *special* characteristics.

In a sense Sraffa added to the confusion rather than solving the problem of combining increasing returns and the theory of market competition. The error Sraffa made was that he did not distinguish between price and marginal revenue, which was remarkable because the concept of marginal revenue had already been developed in a mathematical appendix in Marshall's *Principles*, in which he restates the monopoly theory developed by Cournot.⁶ This was pointed out (again) by Harrod in

⁶ Marshall casts his analysis in terms of net revenue, and only implicitly discusses marginal revenue. The concept of marginal revenue had to be re-invented (Robinson, 1933). This

1930.⁷ For Marshall it was a minor issue and he did not make use of this instrument any further, because he did not need it in his analysis of perfect competition.

This was broadly speaking the state of affairs in the 1920s and 1930s. It was realised that the existence of economies of scale (of one sort or another) implied imperfect market forms, but it remained difficult to construct a satisfactory equilibrium concept for such imperfect market forms. On the one hand there was perfect competition, and on the other hand there was monopoly. Other market forms were considered to be some kind of hybrid of these two extreme forms of competition. So, one could suffice to analyse the two extreme cases in treating all other forms as an implicit mix of the two fundamental forms of competition. But no satisfactory theory of the market existed in which constant or declining marginal and average costs could be made consistent with market equilibrium. This led in the 1930s to a new theory of price determination. One can agree with Schumpeter (1954, p. 1150) that the confusion caused by Marshall was a very fertile one.⁸ Marshall's analysis of the firm and economies of scale led him to develop the concept of the representative firm which invited a lively discussion on market equilibrium and returns to scale and this set the stage for the analysis of monopolistic competition.

1.3 Monopolistic competition in the 1930s

In 1933 two books appeared that changed the way economists dealt with imperfect competition, namely Joan Robinson's *The Economics of Imperfect Competition* and Edward Hastings Chamberlin's *The Theory of Monopolistic Competition*. Although Robinson revived the marginal revolution, in general Chamberlin is considered to be 'the true revolutionary' (Blaug,

is even more surprising considering that Cournot already used the concept of marginal revenue in 1838, and derived the familiar first-order condition for profit maximisation: marginal revenue equals marginal cost (Cournot, 1838).

⁷ See Harrod (1967) for a review of his thoughts on this matter.

⁸ Chamberlin, for example, attributed the origins and inspiration of his theory to the famous Taussig-Pigou controversy on railway rates which took place around 1900. This controversy was about the explanation of different railway rates. Taussig tried to fit different railway rates into the Marshallian theory of (competitive) joint supply by assuming that a unit rail supply is not homogeneous and that different demand elasticities for different stretches of railway result in different prices. In contrast, Pigou stated that it was not an issue of heterogeneity, but of monopoly coupled with the conditions necessary for price discrimination which could explain price differences. In general it is thought that Pigou won the debate.

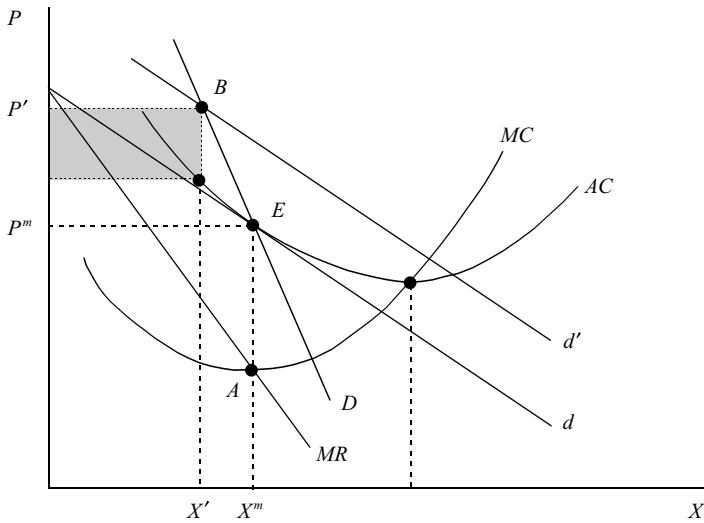


Figure 1.1 Chamberlinian monopolistic competition equilibrium

1997, p. 376).⁹ This radical new analysis was a first answer to the question that was raised in 1926 by Sraffa: is it possible in a market characterised by monopolistic competition and declining average and marginal costs to reach an equilibrium? Figure 1.1 illustrates the equilibrium in the monopolistic equilibrium. Chamberlin makes four basic assumptions (Bishop, 1967, p. 252):

- The number of sellers in a group of firms is sufficiently large so that each firm takes the behaviour of other firms in the group as given (Cournot–Nash assumption)
- The group is well defined and small relative to the economy
- Products are physically similar but economically differentiated: buyers have preferences for all types of products
- There is free entry and exit.

The monopolistic elements are all those elements that distinguish a product from another product and give the firm some market power; ‘each “product” is rendered unique by the individuality of the establishment in which it is sold, including its location (as well as by trade marks, qualitative differences, etc); this is its monopolistic aspect’ (Chamberlin, 1933, p. 63). The large number of firms in the market and the possibility of

⁹ Moreover, the history of Chamberlin’s seminal work dates back to 1921 – see the remarks by Schumpeter (1954, p. 1150).

entry and exit of many firms provides the competitive elements; 'Each [product] is subject to the competition of other "products" sold under different circumstances and at other locations; this is its competitive aspect' (1933, p. 63).

We illustrate the Chamberlinian model with the aid of figure 1.1.¹⁰ We assume that all actual and potential suppliers in the group face the same demand and cost conditions and depict the situation for one particular firm in isolation. There are two demand curves in the diagram. The individual firm under consideration faces demand curve d . This curve represents the firm's price-sales combinations under the assumption that all other firms in the group keep their prices unchanged. Archibald calls this the 'perceived' demand curve (1987, p. 532). The steeper curve labelled D is the demand facing each firm if all firms in the group set their prices identically. Archibald (1987, p. 532) refers to this curve as the 'share-of-the-market' demand curve. As usual MR is marginal revenue (associated with the perceived demand curve d), AC is the firm's average cost, MC is marginal cost, P is the price of the differentiated commodity, and X is the volume of sales.

The Chamberlinian equilibrium under free entry/exit of firms is at point E , where the price is P^m and output is X^m . Point E is the equilibrium because (a) the individual firm attains an optimum in that point, and (b) there are no unexploited profit opportunities, excess profits are exactly zero and no entry/exit of firms takes place. The validity of these requirements can be demonstrated as follows. The individual firm maximises its profit, taking as given the demand curve d . It finds the optimum point by equating marginal revenue and marginal cost (see point A directly below point E). In point E the demand curve d is tangent to the average cost curve, AC , so the firm makes zero profits. This is the famous Chamberlinian tangency condition. Since all firms are identical, no firm makes profits or losses and there is no entry or exit of firms.

Chamberlin (1933, p. 91) also sketched the adjustment process towards the equilibrium point. Assume that all firms in the group are initially operating along the demand curve d' at point B , set a price of P' , and produce a quantity X' . At this price-output combination, each firm would make a positive profit equal to the shaded area in figure 1.1. But point B cannot be an equilibrium. Indeed, in that point the individual firm will have an incentive to lower its price (and increase its profits) by moving to the right along the d' curve (recall that each firm operates under the assumption that its competitors will continue to charge P'). But each firm has exactly the same incentives, so they will all follow suit and cut their

¹⁰ This diagram is adjusted from Bishop (1967, p. 252).

prices. As a result the d' curve will shift down along the D curve towards the Chamberlinian equilibrium at point E .¹¹

Obviously, owing to the downward sloping individual demand curve, there is a difference between equilibrium average cost and minimum average cost in the Chamberlinian equilibrium. This implies that there are unexploited economies of scale and the question arises whether this represents a waste of resources. The answer to this question is both 'yes' and 'no'. 'Yes', in the sense that indeed there is excess capacity and 'no', in the sense that product differentiation introduces variety and this expands the extent of consumer choices and thereby welfare. As Eaton and Lipsey put it, 'in a society that values diversity, there is a trade-off between economizing on resources, by reducing the costs of producing existing products, and satisfying the desire for diversity, by increasing the number of products' (1989, p. 763). We will return to this topic in more detail when discussing the second monopolistic competition revolution.

Given the elegance of the monopolistic competition model it is surprising to see how little influence it had on economic theory. The first attacks on the early monopolistic competition revolution came from Hicks (1939, pp. 83–5) and somewhat later from Stigler (1949) and Friedman (1953). Hicks rejected the theory because he was unable to translate it into a workable model. Stigler (1949) rejected the theory for methodological reasons. He claimed that the predictions derived from the theory of monopolistic competition are not very different from those of perfect competition. Occam's razor then suggests that perfect competition should be favoured over monopolistic competition, a line of reasoning to which Friedman also adheres. It was put forward even more strongly by Archibald (1961, p. 14): 'The theory is not totally empty, but very nearly so' (see also Samuelson, 1967, for a further discussion of this debate). In addition Stigler raised an important point by noting that:

Professor Chamberlin's failure to construct an analytical system capable of dealing informatively with his picture of reality is not hard to explain. The fundamental fact is that, although Chamberlin could throw off the shackles of Marshall's view of economic life, he could not throw off the shackles of Marshall's view of economic analysis. Marshall's technique was appropriate to the problem set to it: it deals informatively and with tolerable logic with the world of competitive industries and monopolies. But it is lost in the sea of diversity and unsystematism, and Chamberlin is lost with it. (Stigler, 1949, p. 22)

¹¹ Note that the position of the D curve depends on the number of firms in the group. In figure 1.1, D is consistent with the Chamberlinian equilibrium at E . As a result, the thought experiment conducted above does not prompt entry of firms. It just shows that E is the only conceivable Chamberlinian (Cournot–Nash) equilibrium.

Archibald (1987, p. 532) mentions two further criticisms that were raised against the Chamberlinian model. First, the notion of the ‘group’ (of products) was ill-defined. In the common definition, goods belong to a group if (1) the cross-elasticity of demand between these goods is ‘high’ and (2) the cross-elasticity between goods in the group and all other goods is ‘low’. The problem with this definition is that there is no logical way to determine what is a high elasticity and what is a low one. Second, Kaldor (1934, 1935) suggested very early on that reality may be better approximated by a market structure with chains of overlapping oligopolies (localised competition) than by Chamberlin’s monopolistically competitive structure. Of course, in such an oligopolistic setting the Cournot–Nash assumption is clearly untenable.

Not surprisingly, in well-known textbooks that appeared in the 1960s and 1970s, monopolistic competition is only briefly mentioned, if at all – see, for example, Henderson and Quandt (1971) and Malinvaud (1972). Akerlof (2002, p. 413) recalls about this period that, ‘monopolistic competition and Joan Robinson’s equivalent were taught in graduate and even undergraduate courses. However, such “specific” models . . . were presented not as central sights, but instead as excursions into the countryside, for the adventurous or those with an extra day to spare’.

The *Festschrift* that was published in honour of Chamberlin also paints a rather bleak picture. Harry Johnson, for example, not only observes that the theory had by 1967 no discernible impact on the theory of international trade, but continues that ‘some beginnings have been made towards the analytical and empirical application of monopolistic competition concepts; but the work has been very much ad hoc, and much synthesizing remains to be done’ (1967, p. 218). What is needed is an ‘operationally relevant analytical tool capable of facilitating the quantification of those aspects of real-life competition’ (1967, p. 218).

But not only Johnson is rather sceptical on the contribution of monopolistic competition; other contributors seem to have the same opinion. Fellner, for instance, concludes that these models are convenient tools of exposition ‘on specific symmetry assumptions . . . In situations lacking these traits of symmetry . . . [they] lose much of their usefulness’ (1967, p. 29) and Tinbergen (1967) observes that the influence on econometrics and macroeconomics is limited.

Only Paul Samuelson is more positive, though still on the defensive, as the following rather lengthy quotation shows:

If the real world displays the variety of behaviour that the Chamberlin–Robinson models permit – and I believe the Chicago writers are simply wrong in denying that these important empirical deviations exist – then reality will falsify *many* of the important qualitative and quantitative *predictions* of the competitive model.

Hence, by the pragmatic test of prediction adequacy, the perfect-competition model fails to be an adequate approximation . . . The fact that the Chamberlin–Robinson model is ‘empty’ in the sense of ruling out few empirical configurations and being capable of providing only formalistic descriptions, is not the slightest reason for abandoning it in favor of a ‘full’ model of the competitive type *if reality is similarly* ‘empty’ and ‘non-full’. (1967, p. 108n, emphasis in the original)

Samuelson concludes that ‘Chamberlin, Sraffa, Robinson, and their contemporaries have led economists into a new land from which their critics will never evict us’ (1967, p. 138).

It might have come as a surprise, even to a relative optimist like Paul Samuelson, that the theory of monopolistic competition was given a new lease on life so quickly. Indeed, less than a decade after the 1967 Chamberlin festivities, Dixit and Stiglitz (1977) managed to again place monopolistic competition theory on the centre stage.

1.4 The second monopolistic revolution

As we pointed out in section 1.3, the monopolistic competition revolution by no means started with the seminal article by Dixit and Stiglitz (1977), but had already had a long (and somewhat troublesome) history. However, one of the reasons why we have gathered the collection of studies in the present volume is that we claim that the second monopolistic competition revolution has been much more successful than the first. The reason for this success is that Dixit and Stiglitz managed to formulate a canonical model of Chamberlinian monopolistic competition which is both easy to use and captures the key aspects of Chamberlin’s model. Though it is by now widely recognised that the Dixit–Stiglitz approach is somewhat unrealistic, it has nevertheless become the ‘workhorse model’ incorporating monopolistic competition, increasing returns to scale and endogenous product variety. As is stressed by Peter Neary in chapter 8 in this volume, the main contributions of the Dixit–Stiglitz model are:¹²

- The definition of an industry (or large group of firms) is simplified: all product varieties are symmetric and are combined in a constant-elasticity-of-substitution (CES) aggregation function (see below).
- Overall utility is separable and homothetic¹³ in its arguments, implying that we can use a two-stage budgeting procedure. In the first stage

¹² There are actually two models in the original Dixit–Stiglitz (1977) paper, which they label, respectively, the Constant Elasticity Case and the Variable Elasticity Case. The first model has become known as *the* Dixit–Stiglitz model. Note that both models have been used in international trade theory, notably Krugman (1979, 1980).

¹³ This is the main distinction from the model developed by Spence (1976), who uses a quasi-linear utility specification.

usually a Cobb–Douglas specification is used, and in the second stage a CES utility function is applied.

- On the production side, technology features increasing returns to scale at firm level. The typical formulation models the average cost curve as a rectangular hyperbola. All firms are symmetrical.

In the remainder of this section we present a very simple version of the Dixit–Stiglitz model and characterise its key properties. Readers who are familiar with the model may skip this section and proceed directly to section 1.5 below.

1.4.1 The model

Preferences

There are two sectors in the economy. The first sector produces a homogeneous good under constant returns to scale and features perfect competition. The second sector consists of a large group of monopolistically competitive firms who produce under increasing returns to scale at firm level. The utility function of the representative household¹⁴ is Cobb–Douglas:

$$U = Z^\delta Y^{1-\delta}, \quad 0 < \delta < 1, \quad (1.1)$$

where U is utility, Z is consumption of the homogeneous good and Y is the consumption of a *composite* differentiated good. This composite good consists of a bundle of closely related product ‘varieties’ which are close but imperfect substitutes for each other. Following the crucial insights of Spence (1976) and Dixit and Stiglitz (1977), a convenient formulation is as follows:

$$Y \equiv \left[\sum_{i=1}^N X_i^{(\sigma-1)/\sigma} \right]^{\sigma/(\sigma-1)}, \quad 1 < \sigma \ll \infty, \quad (1.2)$$

where N is the existing number of different varieties, X_i is consumption of variety i and σ is the Allen–Uzawa cross-partial elasticity of substitution. Intuitively, the higher is σ , the better substitutes the varieties are for each other.¹⁵ In this formulation, $1/(\sigma - 1)$ captures the notion of ‘preference for diversity (PFD)’ (or ‘love of variety’) according to which households prefer to spread a certain amount of production over N differentiated

¹⁴ There is a large number of identical households. To avoid cluttering the notation, however, we normalise the number of households to unity.

¹⁵ In the limiting case, as σ approaches infinity, the varieties are *perfect* substitutes, i.e. they are identical goods from the perspective of the representative household.

goods rather than concentrating it on a single variety (see Bénassy, 1996, for this definition).¹⁶

The household faces the following budget constraint:

$$\sum_{i=1}^N P_i X_i + P_Z Z = I, \quad (1.3)$$

where P_i is the price of variety i , P_Z is the price of the homogeneous good and I is household income (see below).

The household chooses Z and X_i (for $i = 1, \dots, N$) in order to maximise utility (1.1), subject to the definition of composite consumption (1.2) and the budget constraint (1.3), and taking as given the goods prices and its income. By using the convenient trick of two-stage budgeting we obtain the following solutions:¹⁷

$$P_Z Z = \delta I, \quad (1.4)$$

$$P_Y Y = (1 - \delta) I, \quad (1.5)$$

$$X_i = (1 - \delta) \left(\frac{P_i}{P_Y} \right)^{-\sigma} \left(\frac{I}{P_Y} \right), \quad (i = 1, \dots, N), \quad (1.6)$$

where P_Y is the *true price index* of the composite consumption good Y :

$$P_Y \equiv \left[\sum_{i=1}^N P_i^{1-\sigma} \right]^{1/(1-\sigma)}. \quad (1.7)$$

Intuitively, P_Y represents the price of one unit of Y given that the quantities of all varieties are chosen in an optimal (utility-maximising) fashion

¹⁶ In formal terms *average* PFD can be computed by comparing the value of composite consumption (Y) obtained if N varieties and X/N units per variety are chosen with the value of Y if X units of a single variety are chosen ($N = 1$):

$$\text{Average PFD} \equiv \frac{Y(X/N, X/N, \dots, X/N)}{Y(X, 0, \dots, 0)} = N^{1/(\sigma-1)}. \quad (a)$$

The elasticity of this function with respect to the number of varieties represents the *marginal* taste for additional variety which plays an important role in the monopolistic competition model. By using (a) we obtain the expression for the marginal preference for diversity (MPFD):

$$\text{MPFD} = \frac{1}{\sigma - 1}. \quad (b)$$

¹⁷ For a pedestrian derivation of such expressions, see for example Brakman, Garretsen and van Marrewijk (2001, ch. 3).

by the household.¹⁸ Equations (1.4)–(1.5) feature the usual result that income spending shares on Z and Y are constant for the Cobb–Douglas utility function. Equation (1.6) is the demand curve facing the producer of variety i . It features a constant price elasticity, i.e.¹⁹

$$-\frac{\partial X_i}{\partial P_i} \frac{P_i}{X_i} = \sigma.$$

Note that (1.6) provides a formal definition for the individual firm's perceived demand curve (i.e. the d curve in figure 1.1). To derive the industry demand curve (the D curve) we postulate symmetry (see below), set $P_i = P$ and $X_i = X$ (for all $i = 1, \dots, N$), and obtain from (1.6):

$$X = \frac{1}{N} (1 - \delta) \frac{I}{P}. \quad (1.8)$$

Whereas the d curve features a price elasticity of σ (which exceeds unity by assumption), the Cobb–Douglas specification ensures that the D curve is unit elastic, i.e. the *industry* demand curve is less elastic than the demand curve facing individual firms, as was asserted by Chamberlin (1933), and illustrated in figure 1.1, where the D curve intersects the d curve from above.

Technology and pricing

The supply side of the economy is as follows. There is one factor of production, labour, which is perfectly mobile across sectors and across firms in the monopolistic sector. As a result, there is a single wage rate which we denote by W . Production in the homogeneous goods sector features constant returns to scale and technology is given by:

$$Z = \frac{L_Z}{k_Z}, \quad (1.9)$$

where L_Z is the amount of labour used in the Z -sector and k_Z is the (exogenous) technology index in that sector. The Z -sector operates under perfect competition and marginal cost pricing ensures that there are zero

¹⁸ Formally, P_Y is defined as follows:

$$P_Y = \left\{ \min \sum_{i=1}^N P_i X_i \text{ subject to } \left[\sum_{i=1}^N X_i^{(\sigma-1)/\sigma} \right]^{\sigma/(\sigma-1)} = 1 \right\}.$$

¹⁹ In deriving this elasticity, we follow Dixit and Stiglitz (1977) by ignoring the effect of P_i on the price index P_Y . See Yang and Heijdra (1993), Dixit and Stiglitz (1993) and d'Aspremont, Dos Santos Ferreira and Gérard-Varet (1996) for a further discussion of this point.

profits and the price is set according to:

$$P_Z = k_Z W. \quad (1.10)$$

Production in the monopolistically competitive Y -sector is characterised by internal economies of scale. Each individual firm i uses labour to produce its product variety and faces the following technology:

$$X_i = \begin{cases} 0 & \text{if } L_i \leq F \\ (1/k_Y)[L_i - F] & \text{if } L_i \geq F \end{cases}, \quad (1.11)$$

where X_i is the marketable output of firm i , L_i is labour used by the firm, F is fixed cost in terms of units of labour and k_Y is the (constant) marginal labour requirement. The formulation captures the notion that the firm must expend a minimum amount of labour ('overhead labour') before it can produce any output at all (see Mankiw, 1988, p. 9). As a result, there are increasing returns to scale at firm level as average cost declines with output.²⁰

The profit of firm i is denoted by Π_i and equals revenue minus total (labour) costs:

$$\Pi_i \equiv P_i X_i - W[k_Y X_i + F]. \quad (1.12)$$

The firm chooses its output in order to maximise profit (1.12) subject to its price-elastic demand curve (1.6), ignoring the effects its decisions may have on P_Y and/or I (see n. 19). The first-order condition for this optimisation problem yields the familiar markup pricing rule:

$$P_i = \mu W k_Y, \quad \mu \equiv \frac{\sigma}{\sigma - 1}, \quad (1.13)$$

where μ (>1) is the gross markup of price over marginal cost.

Chamberlinian equilibrium

The key thing to note is that the model is completely symmetric. According to (1.13), all active firms face the same price elasticity (and thus adopt the same markup), pay the same wage rate, and face the same technology. Hence, all firms set the same price, i.e. $P_i = P$ for all i . But this means, by (1.6) and (1.11)–(1.12), that output, labour demand and

²⁰ Note that (1.11) implies that the average cost curve of active firms is a hyperbola. This is standard in the Dixit–Stiglitz model. Most graphical presentations of the Chamberlinian model use U-shaped average cost curves. Dixon and Lawler (1996, p. 223) propose the following technology which features a U-shaped average cost curve:

$$X_i = \begin{cases} 0 & \text{if } L_i \leq F \\ (1/k_Y)[L_i - F]^\gamma & \text{if } L_i \geq F \end{cases}$$

with $0 < \gamma < 1$.

the level of profit are also the same for all firms in the differentiated sector, i.e. $X_i = X$, $L_i = L$, and $\pi_i = \pi$ for all $i = 1, \dots, N$. The symmetry property allows us to suppress the i -index from here on.

Before characterising the model developed in this section, we must tie up some loose ends. First, the representative household inelastically supplies H units of labour and is the owner of all firms and thus receives all profits (if there are any). Household income is thus given by:

$$I = HW + N\Pi. \quad (1.14)$$

The second loose end concerns the labour market clearing condition, according to which the demand for labour by the two sectors must equal the exogenously given supply:

$$NL + L_Z = H. \quad (1.15)$$

Owing to its simple structure, the model can be solved in closed form. We start by noting that (1.12) and (1.13) can be combined to obtain a simple expression for profit per active firm in the monopolistic sector:

$$\Pi = W[(\mu - 1)k_Y X - F]. \quad (1.16)$$

With free entry/exit of firms, profits are driven down to zero and the unique output level per active firm follows directly from (1.16):

$$X = \frac{F}{(\mu - 1)k_Y}. \quad (1.17)$$

Output per firm is constant and depends only on features of the technology (F and k_Y) and on the gross markup ($\mu \equiv \sigma/(\sigma - 1)$). The lower is σ , the higher is μ and the smaller is each firm's output. In terms of figure 1.1, the Chamberlinian equilibrium is represented by point E: P^m is given by (1.13) and X^m corresponds to (1.17).

Since profits are zero in the Chamberlinian equilibrium, it follows from (1.14) that $I = HW$ and from (1.4) that $Z = \delta HW/P_Z$. By using this result in (1.9) and (1.10) we find the equilibrium levels of output and employment in the homogeneous goods sector:

$$Z = \frac{L_Z}{k_Z} = \frac{\delta H}{k_Z}. \quad (1.18)$$

A constant share of the labour force is employed in the homogeneous goods sector.

From (1.11) and (1.17) we find that in the symmetric equilibrium $L = k_Y X + F = \mu k_Y X$ or in aggregate terms $NL = \mu k_Y NX$. By using (1.15) and (1.18) we find that $NL = (1 - \delta)H$. Since output per firm is known, we can combine these two expressions for NL and solve for the

equilibrium number of firms:

$$N = \frac{(1 - \delta) H}{\sigma F}, \quad (1.19)$$

where we have used the fact that $\mu \equiv \sigma/(\sigma - 1)$ to simplify the expression somewhat. The equilibrium number of firms depends positively on the amount of labour attracted into the monopolistically competitive sector and negatively on the demand elasticity and the level of fixed cost that each firm must incur. All these effects are intuitive.

Aggregate output of the monopolistically competitive sector can be computed as follows. Equation (1.2) implies that in the symmetric equilibrium $Y = N^\mu X$. By using this result and noting (1.17) and (1.19) we find:

$$Y = \Omega_0 L_Y^\mu, \quad (1.20)$$

where $\Omega_0 \equiv (\sigma - 1)\sigma^{-\mu} F^{1-\mu}/k_Y$ is a positive constant and $L_Y \equiv (1 - \delta)H$ is the total labour force employed in the monopolistically competitive sector. The key thing to note about (1.20) is that, since $\mu > 1$, labour features increasing returns to scale in the Chamberlinian model. Inspection of (1.17) and (1.19) reveals that a larger market (prompted, say, by an increase in the labour force H) leaves the equilibrium firm size unchanged but expands the number of product varieties. Note that by using (1.7) in the symmetric equilibrium, (1.10), and (1.13) we find that the relative price of the composite differentiated good can be written as follows:

$$\frac{P_Y}{P_Z} = \left(\frac{\mu k_Y}{k_Z} \right) N^{1-\mu}. \quad (1.21)$$

This expression provides yet another demonstration of the scale economies that exist in the Chamberlinian model. These scale economies originate from the love-of-variety effect (see also n. 18). Provided μ exceeds unity, the relative price of the differentiated good falls as the number of product varieties rises.

An attractive feature of the Dixit–Stiglitz model is that it contains the perfectly competitive case as a special case. Indeed, by letting σ approach infinity and, at the same time, letting F go to zero, both sectors in the economy are perfectly competitive. Since $\mu = 1$ in that case, it follows from (1.16)–(1.21) that profits are identically equal to zero ($\Pi = 0$), output per firm and the number of firms are undetermined, aggregate output features constant returns to scale and the relative price only depends on relative productivity (k_Y/k_Z).

Welfare properties

Does the Chamberlinian market equilibrium provide too much or too little variety? This is one of the classic questions that has been studied extensively in the monopolistic competition literature. The problem is illustrated by figure 1.1. At point *E* there are unexploited economies of scale owing to markup pricing. Salop (1979, p. 152) uses a spatial model of monopolistic competition and concludes that the market produces too much variety. He is careful to note, however, that this result is not robust. In contrast, in the standard Dixit–Stiglitz model the first-best (‘unconstrained’) social optimum calls for more product varieties than are provided by the market (1977, p. 302) – see also below.²¹ Spence reaches the same conclusion in a special case of his model but argues that the problem is inherently difficult to study because:

there are conflicting forces at work in respect to the number or variety of products. Because of setup costs, revenues may fail to cover the costs of a socially desirable product. As a result, some products may be produced at a loss at an optimum. This is a force tending towards too few products. On the other hand, there are forces tending toward too many products. First, because firms hold back output and keep price above marginal cost, they leave more room for entry than would marginal cost pricing. Second, when a firm enters with a new product, it adds its own consumer and producer surplus to the total surplus, but it also cuts into the profits of the existing firms. If the cross elasticities of demand are high, the dominant effect may be the second one. In this case entry does not increase the size of the pie much; it just divides it into more pieces. Thus, in the presence of high cross elasticities of demand, there is a tendency toward too many products. (1976, pp. 230–1)

In the remainder of this sub-section we study what (our version of) the Dixit–Stiglitz model has to say about this issue.²²

First-best social optimum In the first-best social optimum the social planner chooses the combination of Z , Y and N such that the representative household’s utility (1.1) is maximised given the technology (1.9) and (1.11) and the resource constraint (1.15). In the aggregate this problem can be written as:

$$\begin{aligned} \max_{\{Z, Y, N\}} U &= Z^\delta Y^{1-\delta} \text{ subject to:} \\ H &= k_Y N^{1/(1-\sigma)} Y + FN + k_Z Z, \\ N &\geq N^{MIN}, \end{aligned} \quad (1.22)$$

²¹ In the ‘unconstrained’ social optimum, only the resource constraint is taken into account. In the ‘constrained’ social optimum the additional requirement of non-negative profit per active firm is added.

²² The welfare analysis follows the approach of Broer and Heijdra (2001).