Tempora mutantur: In search of a new testament for NEG

Kristian Behrens*  Frédéric Robert-Nicoud†

October 15, 2010

— Revised version —

Abstract

The aim of this article is twofold. First, we briefly discuss the constraints associated with the canonical ‘Cobb-Douglas—Dixit-Stiglitz-Spence—Samuelson’ NEG setup originally used by Krugman in 1991. We outline the key features of his CP model and highlight what results are robust to changes in functional forms and agglomeration mechanisms. We also selectively review the literature that amends the original CP model in various directions and discuss what qualitative insights do change. Our conclusion is that progress within the straightjacket of Krugman’s original framework is unlikely to produce path-breaking new insights. Second, we suggest that there are potentially large payoffs to stepping outside of the established framework and to extend the NEG approach into various directions that have to date received only little attention. Heterogeneity, cities, transportation, public policy and calibration are avenues along which NEG needs to make progress.

Keywords: New Economic Geography; core-periphery model

JEL Classification: R10; R12; R13; R23

*Canada Research Chair, Département des Sciences Économiques, Université du Québec à Montréal (UQAM), Canada; CIRPÉE, Canada; and CEPR, UK. E-mail: behrens.kristian@uqam.ca
†Département d’économie politique, Université de Genève, Switzerland; Spatial Economics Research Centre and Centre for Economic Performance, London School of Economics (LSE), London, UK; and CEPR, UK. E-mail: frederic.robert-nicoud@unige.ch
1 Introduction

Two decades ago, Paul Krugman’s seminal article “Increasing returns and economic geography” (*Journal of Political Economy* 99, 1991, pp.483–499) launched the New Economic Geography (NEG) as a subfield of the broader field that deals with the location of economic activity.\(^1\) Since then, countless refinements of his core-periphery (CP) model have been developed to analyse regional divergence, the formation of industrial clusters, and the impacts of regional policy, to name but a few. NEG has been very successful in providing rationales for the existence and the persistence of sizable regional disparities at several spatial scales. Krugman’s work has achieved to draw spatial economics from the periphery to the center of economic theory, thereby making NEG an important component of the spatial economist’s analytical toolbox.

Despite such “rapid progress [that] would make anyone dizzy” (Neary, 2001, p.536), the increasing-returns paradigm of NEG seems dangerously close to be running into a decreasing-returns phase. Though this is not a feature specific to the NEG, the dearth of NEG papers in top economics journals suggests that its current contributions are no longer perceived as pushing sufficiently the mainstream economic frontier. One potential reason for this state of affairs is that many post-Krugmanian contributions have been incremental: NEG models are quite difficult to work with and, as a result, most authors stick to a relatively narrow framework that closely follows Krugman’s initial setup. This is not to say that such increments are unimportant. The finding whereby the overwhelming share of NEG models — including those built on the quite different framework by Ottaviano *et al.* (2002) — display relatively common equilibrium properties is a fundamental one as it testifies to the robustness of the key insights. However, it also seriously limits the additional knowledge that we may hope to gain from simple variations on NEG models, especially Krugman’s one. New insights can only be obtained by departing from some of the fundamental assumptions and functional forms underlying the original CP model, and by focusing on broader questions for which the spatial dimension has been insufficiently explored to date. In other words, a ‘New Testament’ needs to supplement the current ‘Bible.’

Krugman’s 2008 Prize in Economic Sciences in Memory of Alfred Nobel provides a unique opportunity for NEG to ponder where this New Testament should take us and to establish itself more firmly in the economic mainstream. At a certain level of generality, all economic problems arise because economic agents are heterogenous along some dimension and because this heterogeneity gives rise to gains from trade that they seek to exploit. *Time* and *space* are two dimensions along which such gains may be generated and, at the same time, two types of frictions that impede the realization of these gains. Time is already well integrated into any economist’s toolkit. Space is still remarkably absent from mainstream economics. This is an anomaly, and one that NEG can alleviate if it finds a way to communicate its insights in a parsimonious way.

\(^1\)That trade theory and location theory largely deal with similar issues has been known since Ohlin’s 1933 contribution (see also Krugman, 1993). Fujita and Thisse (2002, 2009) provide a comprehensive exposition of how NEG fits into the broader field that deals with the location of economic activity (including land-use theory, urban economics, location theory and spatial competition). As pointed out by these authors, many concepts and ideas ‘rediscovered’ by Krugman had been around for a while. Yet, they had not been put together and explored in a systematic way until then.
Against this backdrop, the aim of this article is twofold. First, we briefly discuss the constraints associated with the ‘Cobb-Douglas—Dixit-Stiglitz-Spence—Samuelson’ setup originally used by Krugman in 1991. We draw on the key features of the CP model and highlight what results are driven by which assumptions. We also selectively review the literature that amends the original CP model in various directions and discuss what qualitative insights emerge. Our conclusion is that progress within the straightjacket of that framework, though potentially important, is unlikely to produce path-breaking new insights. Second, we suggest that there are potentially large payoffs to stepping outside of the established framework and to extend the NEG approach into various directions that have to date received only little attention. Heterogeneity, cities, transportation, public policy and calibration are among the avenues along which theoretical NEG needs to make progress. Although our list of suggestions will look familiar to most readers, echoing various points that have been emphasized before (e.g., Fujita et al., 1999; Fujita and Mori, 2005; Fujita and Thisse, 2009) these are still, we believe, the key directions into which NEG needs to move as little has been done so far.

Three final comments are in order. First, a paper reaches the ultimate stage of fame when it is referred to in a text without appearing in the reference list. The specification of preferences put forth by Cobb and Douglas, the constant elasticity of substitution (CES) monopolistic competition model due to Spence, Dixit and Stiglitz, and iceberg trade costs due to Paul Samuelson are prime examples. It therefore seems appropriate that Krugman’s founding article suffers the same fate and does not appear in our reference list. We apologise to all authors whose papers are listed in our bibliography. Second, for lack of space, this short article is selective in the topics and papers it covers. This selection reflects both our research interests and our ignorance, and is not a subjective attempt to evaluate the relative importance of the various contributions to the field. Last, this paper deals almost exclusively with theoretical aspects of NEG and deliberately does not discuss the numerous important developments that have taken place on the empirical frontier. The reasons for this are our comparative disadvantage in the matter, lack of space and the existence of numerous excellent recent surveys on the topic to which we have nothing to add (e.g., Head and Mayer, 2004 and 2010; Redding, 2010).

2 A class of NEG models: Games you can play with Cobb-Douglas, Dixit-Stiglitz-Spence, and Samuelson

We start by discussing briefly the most distinctive features of the CP model and how these are shared by a wide class of alternative models. The fundamental result here is that all of Krugman’s positive results but one are robust to the choice of functional forms and agglomeration mechanisms. We also discuss extensions that step outside his canonical framework and stress that some of them are widely influential precisely because they went beyond this canonical framework to bring fresh insights. We then review the normative implications of NEG models; the important result here is that these are very sensible to the choice of functional forms and agglomeration mechanisms, unlike the positive properties.
2.1 Key features and results: ‘New NEG models, same old properties’

As is well-known, Krugman’s seminal CP model has two fundamentally similar regions (points on the map), two sectors, and two factors. Each factor is specific to one sector — one factor is inter-regionally mobile while the other is not. The CES differentiated good uses the mobile factor and is produced under increasing returns to scale (IRS) that are internal to the firm. The homogenous good sector employs the immobile factor under constant returns to scale (CRS). Both goods are tradable — trade in the former is impeded by iceberg transportation costs, whereas trade in the latter is costless. Two types of stable equilibria exist in this model: agglomeration of the CES industry in a single region when transportation costs are low; even dispersion of this industry across the two regions when transportation costs are high. Both equilibria are stable for intermediate values of transportation costs. Though this is a very stylized model, few papers in the field have dared stepping outside this peculiar framework. Many more have established that some of the key and most striking features of Krugman’s model are robust to marginal changes in functional forms and, more importantly, are independent of the agglomeration mechanism — be it labour migration as in the Krugman’s CP model, input-output linkages as in Venables (1996) or (human) capital accumulation as in Baldwin (1999) and Martin and Ottaviano (1999, 2001). All these models share seven famous features that are detailed below. This is remarkable. The key insight is that, provided that goods are produced under IRS (so that firms have a small number of plants) and markets are segmented (so that the plant location decision is not trivial), what happens in one region has direct and indirect effects on the other one. In the case of labour migration, as in the CP and Forslid and Ottaviano (2003) models, this sounds probably obvious: after all, inward migration in one region is the mirror image of outward migration in the other one. But the effects on the economic geography of the two regions if labour is mobile across sectors instead of regional borders (Krugman and Venables, 1995; Ottaviano and Robert-Nicoud, 2006; Robert-Nicoud, 2006) or if (human) capital is accumulated in one region but not the other are similar to those at work in the CP model: migration, labour reallocation towards the CES-industry and human capital accumulation all change the relevant market sizes for CES-good producers. Such changes are at the core of the agglomeration and dispersion forces of NEG models; these then display seven features (see Baldwin et al., 2003, on this theme):

1. **Home market magnification.** Krugman’s (1980) home-market effect (HME), whereby an exogenous shift in demand for the CES good results in the more than proportional re-location of industry, operates if (and only if) markets are segmented and goods are produced under IRS. The HME distinguishes NEG and ‘new trade’ models from neoclassical ones and has been exploited in the empirical literature by Davis and Weinstein (2002, 2008) and Brüllhart and Trionfetti (2009); see Behrens et al. (2009) for a generalization and Head and Mayer (2004) for a review. It turns out that the HME is exacerbated as trade becomes freer.

2. **Circular causality.** Put simply, a region’s CES industry may be big because this region has a high income — and income tends to be high in regions that have a lot of industry. Specifically, agglomeration forces in the CP models and its aforementioned variants are self-enforcing: if the market size of CES goods of a region becomes relatively larger as a result of in-migration, labour
reallocation in the presence of input-output linkages or capital accumulation, then this increases the region’s attractiveness for the location of plants in the CES industry. Likewise, an increase in the relative number of plants in one region tends to lower the local price of the CES composite input and/or consumption basket relative to the other region, thus raising the profitability of the CES industry (or increasing consumers’ utility) in that region relatively to the other one, and hence the cycle repeats. The location equilibrium is not degenerate because, working in the other direction, is the market crowding effect: as more firms cluster in a region, each of them has a smaller market size. Clearly, market crowding is a dispersion force.

3. Endogenous asymmetry. When agglomeration forces dominate dispersion forces, the symmetric equilibrium becomes unstable and one region grows at the expense of the other.

4. Catastrophic agglomeration. This feature is shared by all the simplest models with Cobb-Douglas–cum–CES preferences, Samuleson iceberg costs and constant returns to labour. Yet, it is not robust to small perturbations of the model (Berliant and Kung, 2009).² It implies that small, continuous variations in parameter values have discrete ‘bang-bang’ effects on the equilibrium location of industry.

5. Overlap and self-fulfilling expectations. For some parameters of the model, multiple equilibria exist. In such situations, history and expectations matter and may help in selecting the equilibrium (Matsuyama, 1991; Krugman, 1991; Baldwin, 2001; Oyama, 2009).

6. Locational hysteresis. This feature is closely related to the previous one. To see the distinction, consider myopic agents and a reduction in transportation costs below the so-called break point: the stability of the symmetric equilibrium is thus ‘broken’ and the CES sector clusters in a single region, which becomes the ‘core’. If transportation costs then rise again slightly above the break point, the CES industry stays put in the core. Thus, temporary policy changes may have permanent effects. This property of NEG models has been empirically tested by Davis and Weinstein (2002), Bosker et al. (2009), Redding et al. (2009) and others.

7. Hump-shaped agglomeration rents. Location is lumpy and exhibits hysteresis because there are agglomeration rents: at the core-periphery equilibrium, firms are generally not indifferent at the margin as to where to locate. These rents are highest for intermediate values of transportation costs. Indeed, these costs need to be low enough to make agglomeration sustainable and high enough to make a real difference in profits. These rents can be taxed without triggering capital outflows or de-industrialisation in the core. Tax competition may even trigger a ‘race to the top’ (Baldwin and Krugman, 2004; Kind et al., 2000; Ludema and Wooton, 2000).

The fact that all simple NEG models with Cobb-Douglas–cum–CES preferences, Samuleson iceberg costs and constant returns to labour share these features is, as we said, remarkable. It turns out

²As an illustration, consider Pflüger (2004) who uses quasi-linear preferences instead of Cobb-Douglas. This seemingly innocuous change is enough to rule out catastrophic agglomeration.
that these models are even *isomorphic* in an economically meaningful state space (Robert-Nicoud, 2005; Ottaviano and Robert-Nicoud, 2006; Behrens and Robert-Nicoud, 2010b). Therefore, it is our contention that not much new can be learned by playing with marginal variations on this canonical NEG framework. More substantial departures are required.

Some of the most celebrated developments in the field have precisely stepped out of the canonical framework. Ottaviano *et al.* (2002), for example, use an entirely different demand structure. In this alternative monopolistic competition framework, markups are monotonically decreasing in the number of local competitors — a highly desirable and realistic feature. Another distinction with Dixit-Stiglitz-Spence's monopolistic competition framework is that this one has no income effect. Yet, this model still displays 5 of the famous-7 features of the canonical NEG framework\(^3\) This suggests that most of these features are indeed quite robust to the choice of functional forms. Another widely celebrated development in the field is the bell-shaped curve of development uncovered in Krugman and Venables (1995) and extended in Puga (1999). These authors add decreasing returns to labour in the traditional sector to an otherwise canonical NEG model with input-output linkages à-la Venables (1996). The regional labour supply curve to the CES industry is thus upward-sloping in each region. This adds a dispersion force to the model as labour costs rise when a region industrialises. This dispersion force is unrelated to the level of transportation costs, thus dispersion of industry is the unique equilibrium outcome when transportation costs are arbitrarily low. Also, the transition from dispersion to agglomeration and dispersion again as transportation costs fall is smooth — no catastrophic agglomeration here. Rather, a wave of industrialisation of the North is followed by a wave of re-industrialisation of the South. This model is thus one plausible explanation for the de-industrialisation of India and China in the nineteenth century and the current re-industrialisation of these countries and others (Puga and Venables, 1996). Extensions of this model have also been used to study the impact of offshoring on these opposite industrialisation experiences (Fujita and Thisse, 2006; Robert-Nicoud, 2008). Finally, the introduction of housing markets, urban congestion and competition for land, all of which are empirically important, have been shown to alter the equilibrium patterns even more dramatically (Helpman, 1998; Murata and Thisse, 2005; Pflüger and Tabuchi, 2010).

In hindsight, it is thus fair to say that catastrophic agglomeration is the least robust property of NEG models. It also seems to be about the only artifact in this literature. The cause of its existence lies with two, somewhat implicit, key assumptions: people are homogenous and mobility is frictionless. As we shall develop in Section 3, relaxing the first assumption in particular brings more plausible results and suggests some interesting avenues for further research.

### 2.2 Normative properties and welfare implications

Existing positive differences within the canonical NEG framework are by now relatively well understood and rather modest. By contrast, the differences in the normative properties of NEG models are less well-

\(^3\)The model of Ottaviano *et al.* (2002) has no overlap and no locational hysteresis. A further robustness check of the NEG results could be to apply the new monopolistic competition model of Behrens and Murata (2007), which has both income and pro-competitive effects, to these questions.
known and more substantial. Charlot et al. (2006) show that agglomeration and dispersion cannot be Pareto-ranked in Krugman’s original model. Furthermore, except when transport costs are sufficiently low, no equilibrium is preferred to another based on compensation criteria à la Kaldor and Hicks. Using social welfare functions of the generalised-mean type allows one to reach sharper conclusions which, however, “heavily [depend] on societal values regarding inequalities across individuals” (Charlot et al., 2006, p.325). The latter being unobservable, debates about whether the market yields too much or too little agglomeration may not be very illuminating in the end. The quadratic-linear model developed by Ottaviano et al. (2002) allows for a much simpler utilitarian welfare evaluation because of constant marginal utility of income due to quasi-linear preferences. Ottaviano and Thisse (2002) show that agglomeration is both the optimum and the equilibrium outcome for low values of trade costs, whereas dispersion is both the optimum and the equilibrium outcome for high values of trade costs. By contrast, the market provides too much agglomeration for intermediate values of trade costs.

Though there seems to be some (predominantly political) consensus that the market may indeed generate too much agglomeration, one should keep in mind that this finding needs to be qualified along several lines. Firstly, NEG welfare results do crucially hinge on whether or not the analysis includes additional ‘agglomeration economies’ which are empirically important (see Rosenthal and Strange, 2004; Kerr et al., 2010; Melo et al., 2010; Puga, 2010). Ottaviano and Robert-Nicoud (2006) establish that agglomeration economies in the form of input-output linkages can be strong enough for agglomeration to Pareto-dominate dispersion: in that case, agglomeration also benefits those left behind in the periphery via lower prices and richer product diversity. By the same token, insofar as agglomeration of economic activity may sustain overall growth, per-capita incomes may also be higher in the periphery under agglomeration than under dispersion (Baldwin et al., 2001). Secondly, NEG models assume that mobile factors are perfectly mobile. However, in the presence of impediments to mobility, it may well be that too little agglomeration arises at equilibrium, which can be costly in terms of output and growth (Au and Henderson, 2006). Finally, welfare results also depend heavily on the nature of the dispersions forces involved. As shown for example by Helpman (1998), the market outcome may well yield too little agglomeration in the presence of urban costs, even within an otherwise relatively standard CP model.

3 In search for a new testament: Heterogeneity, cities, transportation, public policy and calibration

A decade ago, Fujita et al. (1999, p.6) subsumed the NEG state-of-the-art with the slogan “Dixit-Stiglitz, Icebergs, Evolution, and the Computer”. A decade later, not that much has changed despite repeated calls and various research agendas (e.g., Neary, 2001; Fujita and Mori, 2005). Dixit-Stiglitz remains the workhorse, though several alternative frameworks have allowed us to broaden our under-

---

4Gaigné (2006) shows that the Pareto criterion is sensitive to whether the core is fully specialised in the monopolistically competitive sector or not. When this is so, dispersion Pareto-dominates agglomeration as rising wages in the core reduce the benefit of lower prices to the periphery.
standing of agglomeration phenomena. *Icebergs* are omnipresent — no global warming here — though it should be obvious that many results in both trade and Neg do hinge on this peculiar assumption. *Evolution* is a key technical trick to deal with multiple equilibria and catastrophic change which, depending on your perspective, are either a desirable key feature or an undesirable nuisance. And, last but not least, the computer is still widely used to provide ‘insights’ into the models’ behaviors. Though none of these four points is likely to change in the foreseeable future, it is our contention that key breakthroughs can only be made by pushing beyond them.

We now selectively discuss a few directions along which we think theoretical Neg needs to make progress in the future. In so doing, we deliberately skip empirical issues, which would deserve another paper of equal length to the present one. Our choice of menu will look familiar to most readers, echoing various points that have been emphasized before (e.g., Fujita *et al.*, 1999; Fujita and Mori, 2005; Fujita and Thisse, 2009).\(^5\) These are however still, we believe, the key directions into which Neg needs to move. As little has been done thus far it is worth hammering home those points one more time.

**Heterogeneity and the city.** Dixit-Stiglitz monopolistic competition continues to dominate Neg and many other fields due to its analytical convenience. The framework has been enriched with firm heterogeneity which has become an important topic in e.g. macroeconomics and trade (see the article by Ottaviano in this issue). Following the seminal works by Melitz (2003) and Jean (2002), Dixit-Stiglitz monopolistic competition can now cope with heterogeneous producers and, therefore, tell us something about firms, industry dynamics, and selection processes. The same holds true for the quadratic-linear model that has been extended by Melitz and Ottaviano (2008) to cope with firm heterogeneity. Given the tight link between trade and Neg (some may argue that this is a ‘leader-follower’ relationship), heterogeneity has begun to play a limited yet increasing role in analyzing various phenomena like the spatial sorting of firms or the existence of regional productivity differences.

Firm-level productivity differences have some clear implications for the spatial structure of the economy. Baldwin and Okubo (2006) and Nocke (2006), for example, show that more productive firms sort themselves into larger markets, and Okubo and Picard (2008) show that firms producing higher ‘quality’ goods are also attracted to those larger markets. These results suggest that there is complementarity between regional size and firm-level productivity. Behrens and Robert-Nicoud (2010a) push this idea further by developing a multi-region agglomeration model in which the number and the size of cities is endogenously determined (see also Fujita *et al.*, 1999b). Such an urban Neg model with worker heterogeneity can replicate three sets of facts: the size-productivity relationship in an urban hierarchy, as amply documented by empirical studies; the size-income inequality relationship in an urban hierarchy (Glaeser and Maré, 2001; Wheeler, 2001; Baum-Snow and Pavan, 2010); and the existence of episodes of massive urbanisation, whereby new cities quickly appear and existing cities significantly grow in response to small changes in trade costs. The latter point can also rationalise

\(^5\)In the wake of Krugman’s Nobel Prize, several articles have recently reviewed Krugman’s contributions and speculated on future research directions. All these papers have, quite naturally, a significant overlap in the avenues they explore and in the topics they discuss. See, e.g., Fujita and Thisse (2009), Brakman and Garretsen (2009), and Behrens and Robert Nicoud (2009).
the fact that more open economies have higher urbanisation rates and less urban primacy (Ades and Glaeser, 1995). Traditional two-region NEG models are silent on urbanisation and would predict either more ‘agglomeration’ (i.e., urban primacy) if dispersion forces are of the ‘pull’ type, and the opposite if dispersion forces are of the ‘push’ type.

Two well-documented empirical facts that still seem largely out of reach of NEG models are the spatial sorting of heterogeneous individuals across cities and an explanation for an urban hierarchy following the rank-size rule (in the words of Krugman, 1996, p.399: “while there must be a compelling explanation of the astonishing empirical regularity [the power law], I have not found it.”). Behrens et al. (2010) propose a model based on NEG-type input-output linkages in which an urban hierarchy emerges endogeneously. Using a combination of urban system theory à la Henderson (1974), input-output linkages à la Ethier (1982), and firm heterogeneity à la Jean (2002) and Melitz (2003), the model generates a system of cities where heterogeneous individuals sort themselves across space. At equilibrium, the positive size-productivity relationship holds and (under some empirically plausible conditions) the size distribution of cities follows Zipf’s law.

As should be clear from the foregoing, we think that factor heterogeneity and urban ingredients are important for future developments of the NEG. The former allows us to bring the models closer to the empirical facts and makes NEG models more amenable to numerical applications by providing an ‘antidote’ to multiple equilibria (more on this below). The latter bridges the gap with urban economics; this is an important task as most ‘agglomerations’ are in fact large urban areas within which various congestion costs and productivity differences are key features (see also Thisse, 2010). Taking a closer look at urban elements and heterogeneity is also likely to increase the policy relevance of NEG models which, as argued below, is fundamental.

**Beyond the tip of icebergs.** Transportation costs play a central role in the NEG. Yet, paradoxically, the modeling of transport costs is arguably one of its weakest parts. Until now, the key insights of NEG are about how changes in some exogenous parameter \( \tau \) drive the equilibrium structure of the space-economy. Worse, the iceberg assumption is far from being innocuous. While many authors (including in international trade) have complained about the restrictiveness and the potential drawbacks of this assumption (Neary, 2001; Fujita and Thisse, 2002; McCann, 2005; Matsuyama, 2007), little progress has been made in ‘sinking the iceberg’.

It is time to start to seriously think about transportation in NEG. *Transport costs are prices which are set in (mostly) imperfectly competitive markets and which respond to demand and supply shocks.* The recent trade collapse provides a neat illustration: the Baltic Dry Index (BDI) fell between May 20, 2008 and December 5, 2008 by a whopping 94%, and this may have been the burst of the biggest of all bubbles. Technology certainly did not change that dramatically over such a short period —

---

6It is well known that heterogeneity ‘convexifies’ problems and makes them smoother. As Krugman (1993, p.116) already noted, “it is necessary to “fuzz up” the model.” This seems especially important for the future development of calibrated multi-region models.

7The BDI is an index published by the Baltic Exchange that keeps track of international shipping prices for dry bulk cargoes like coal and grain along 20 major routes.
supply and demand did. Future theories of a spatial economy ought to take such aspects into account by endogenizing transport costs (Matsuyama, 2007; Behrens et al., 2009).

Another empirical fact that NEG has to come to grips with is that transport and trade costs are not symmetric. Jonkeren et al. (2010) provide empirical evidence that freight imbalances affect transport costs substantially: a one standard deviation increase in the trade imbalance between two regions decreases transport prices on the short side of the market by about 7%. Behrens and Picard (2008) and Takahashi (2010) show that these freight rate asymmetries arising from trade imbalances have strong effects on the spatial structure of the economy: full agglomeration never occurs in the core-periphery model and the home-market effect is substantially weakened or even reversed.

Last, one additional but less well-known drawback of the iceberg emerges in the presence of firm heterogeneity. Indeed, since the iceberg formally amounts to a scaling of the firm’s marginal cost, it follows that more productive firms face de facto lower per unit trade costs. This has surprising implications. For example, in Melitz and Ottaviano (2008), the most productive firms always serve distant markets through trade, the reason being that shipping is basically costless for the most efficient firms. Hence, in the context of the proximity-concentration tradeoff, only firms with intermediate productivity would choose to serve foreign markets via foreign direct investments (FDI). This result is counterfactual (Helpman et al., 2004) and it could only be reversed if the fixed costs of FDI are also tied to firms’ productivity. To our knowledge, there is no strong empirical evidence supporting the productivity-lower trade cost link. Furthermore, it is also unclear what theoretical mechanisms could generate such a relationship. Since the bulk of manufacturing firms outsource the transportation of their goods (about 80% of US firms in 1992, according to the U.S. Bureau of Transport Statistic), a tentative theoretical explanation for why more productive firms would face lower shipping costs could be assortative matching between productive manufacturing firms and productive carriers. We are aware of no empirical evidence for this either. Finally, since more productive firms sell larger volumes, bulk discounts by carriers who operate under increasing returns to scale offer an additional explanation. Exploring these issues more fully requires, by definition, that the transport sector be explicitly modeled. Maybe the iceberg will turn out not to be that bad an assumption after all, but it would be reassuring to know why and how this short-cut can be used.

Multiple equilibria, evolution and computers. As argued in the foregoing, the analytical properties of the CP model are now well known (Robert-Nicoud, 2005; Mossay, 2006). Fujita and Thisse (2009, p.116) rightfully point out that “we don’t need the computer anymore to study the core-periphery model”. Well, not quite if we want to take the model to data beyond two regions. In such multi-region settings, equilibria are often an embarrassment of riches. While some view this as a desirable key feature of NEG models, others consider it an undesirable nuisance. We take no stand on this issue, but we think it is fair to say that multiple equilibria make our life miserable when it comes

---

8NEG is of course not the only field in which multiple equilibria abound. As pointed out by a referee, this feature is so pervasive that Debreu (1970) wrote a paper with a very suggestive title on that topic: “Economies with a finite number of equilibria”.

10
to more applied work, be it empirical or simulation based. Small changes in parameter values can unravel the simulation exercise, the models behave in a rather unstable way, and empirical work based on comparative statics has to make the leap of faith that the data variations are within ranges for which there are no switches across different equilibrium regimes. The multiplicity of equilibria is a deep feature of NEG models that can, however, be dealt with in several ways. First, we can recognize that it is partly an artifice due to implausibly strong symmetry assumptions which usually lead to multiple mirror-equilibria. The latter can be ruled out by ‘injecting’ some heterogeneity into the model (e.g. Desmet and Rossi-Hansberg, 2009, 2010). Breaking the symmetry then possibly reduces the set of equilibrium configurations or allows at least to focus on more plausible sets of equilibria by making the basins of attraction of some implausible equilibria very small. The downside of breaking the symmetry is, of course, that the computer becomes once more the NEG researcher’s best friend. Second, we can supplement NEG models with an equilibrium selection mechanism beyond the simple dismissal of unstable equilibria or myopic evolutionary dynamics. Planners or land developers, together with permanent investments, could play this role (Henderson and Venables, 2009). Last, multiple equilibria become less of a problem the less mobile agents are. Heterogeneity in terms of regional attachment (Tabuchi and Thisse, 2002; Murata, 2003), moving costs, or local amenities that do not depend on market size will make multiple equilibria less of a problem by reducing the responsiveness of agents to differences in economic conditions across regions. Unfortunately, too little work has been done on these topics until now.

Research in NEG should make intelligent use of the power the computer holds. We need to better understand and design the calibration exercises and quantified models.\(^9\) This will allow us to run counterfactual simulations which are already largely used in macroeconomic models and in international trade (e.g., Eaton and Kortum, 2002; Bernard et al., 2003).\(^10\) NEG is lagging behind and since 1991 little progress has unfortunately been made in this direction. We conjecture that this is one explanation for the underrepresentation of the field in the top journals of the profession. We also conjecture that if NEG is to be taken seriously as a tool that allows us to understand the workings of a spatial economy, we need to put numbers on the theory and to get a good understanding of the magnitudes involved. To this end, serious calibration is required, which is to date rarely (if ever) done. We have had enough simulations with highly implausible parameter values and little to no policy relevance.

**Public policy.** Spatial imbalances at almost any scale preoccupy policymakers (see, e.g., World Bank, 2009). They also motivate research in regional science in general and NEG in particular. Yet,

\(^9\)One of our referees pointed out that, “the only solution is to build estimated and computable general equilibrium models and to proceed to real simulations, something economists have not learnt yet.”

\(^10\)The potential payoffs to plausible and workable multi-region NEG models that allow for counterfactual applications are likely to be high. Many interesting questions could be investigated using such models. For example, one could explore in a much more systematic way the magnitude required for shocks to city size before the urban hierarchy changes its structure. This might provide more structural tests in the wake of Davis and Weinstein (2002) and Bosker et al. (2006). Another application would be to use the models to ‘back out’ unobservables like productivity or ‘amenities’ from the calibration exercise. This would allow for Rosen-Roback type of analyses within a full-fledged NEG general equilibrium framework in which geography, product diversity and scale economies matter.
the latter has almost nothing to say to these policymakers. The lack of sharp policy conclusions has led many authors, including the main proponents of NEG, to refrain from prescribing any policy recommendations. This has left the door open for ‘regional pseudo-scientists’ to perform those policy recommendations, often without using any explicit welfare criterion, microeconomically founded model, or empirical analysis.\textsuperscript{11} NEG researchers need to claim this terrain. It was the contention of Baldwin \textit{et al.} (2003) that the lack of sharp policy implications called for more work, not less. This remains true seven year later. So far, almost nothing has been done except in the area of tax competition (e.g., Baldwin and Krugman, 2004; Kind \textit{et al.}, 2000; Ludema and Wooton, 2000). This is an anomaly that cries for more new research.

The relative dearth of clean empirical work and of credible calibration exercises does not make matters any simpler for NEG in the policy arena — although recent high-quality empirical work by e.g. Redding and Venables (2004), Hanson (2005) or Redding and Sturm (2008) is encouraging (see Head and Mayer, 2004, and Redding, 2010, for surveys). Much remains, however, to be done if NEG is to ever become an important tool for regional policy. As argued above, most current NEG calibration and counterfactual exercises fall short of the standards of the macro literature. Too often, papers rely on \textit{one} simulation using seemingly arbitrary parameter values instead of disciplining their model by trying to replicate some moments in the actual data. For these reasons, it is currently still very difficult to obtain a precise idea of how regional policy would affect the spatial structure of the economy and what the associated productivity and welfare changes could be.

Despite its shortcomings, NEG has had at least two important merits when it comes to public policy. First and foremost, it has put geography at the heart of the question of regional inequalities and development. Only too often have policy makers looked (or still look) at regions as ‘floating islands in space’ that can be analyzed in isolation.\textsuperscript{12} NEG forcefully reminds us that regions are interconnected and cannot be analyzed individually. Second, NEG points to the importance of considering the spatial structure of the economy as being endogenous when thinking about the impacts of policies. This is important as changes in the spatial structure of the economy may significantly amplify or dampen the potential benefits of public investments and policies. Venables (2007), for example, shows that NEG-type agglomeration economies are likely to magnify to benefits from urban transport investments by changing the size of the agglomeration in response to those investments. Behrens \textit{et al.} (2009) show that deregulation of the transport sector may lead to smaller welfare gains when the spatial structure of the economy changes in response to that deregulation than when it is considered as given. Those examples, though derived in simple frameworks, suggest that policy makers need to carefully consider how the spatial structure of the economy might change in response to their policies to fully gauge their potential impacts. This has yet to become standard practice.

\textsuperscript{11}See Duranton (2010) for a critical review of the theoretical foundations and empirical impacts of ‘cluster policies’.
\textsuperscript{12}We thank Steven Brakman for suggesting this point.
4 Conclusions: Always look at the bright side of life

NEG has reached a certain level of conceptual and theoretical maturity, especially in its two-region Dixit-Stiglitz-Iceberg form. Most of the low-hanging fruits of the NEG tree have been harvested by now — despite many ‘new’ models, we often get the same old results. We think, as many did before us, that it is time to step back, take a closer look at the forest behind the tree, and seriously ponder about the way forward. NEG remains as exciting an area of research nowadays as it has ever been. Yet, we must collectively realise that the focus needs to shift towards new questions and away from the status quo. Only by doing so will NEG and spatial economics more generally be able to claim its rightful place as one of the important branches of economics. Should we fail to do so, Neary’s (2001, p.557) gloomy prediction about the future of NEG might well come true: “I suspect that [NEG] will [survive], though maybe not as a distinct field. Instead, I am tempted to suggest that it will survive as ‘merely’ another simple general equilibrium model, supplementing the trade theorist’s tool kit.” We would rather always look at the bright side of life.

Acknowledgements. We thank Steven Brakman, Harry Garretsen and two anonymous referees for very helpful comments and suggestions. Behrens is holder of the Canada Research Chair in Regional Impacts of Globalization. Financial support from the CRC Program of the Social Sciences and Humanities Research Council (SSHRC) of Canada is gratefully acknowledged. Behrens also acknowledges financial support from FQRSC Québec (Grant NP-127178). Robert-Nicoud thanks IES Princeton for its hospitality. “We blame each other for all remaining errors.” (Dixit, A. and G. Grossman, 1982, “Trade and protection with multistage production”, Review of Economic Studies 42, 583-594).

References


