



BANKWEST CURTIN ECONOMICS CENTRE

# WORKING PAPER SERIES

19/04: THE UNINTENDED CONSEQUENCES  
OF INCREASING RETURNS TO SCALE IN  
GEOGRAPHICAL ECONOMICS

Steven Bond-Smith

This report was written by researchers affiliated with the Bankwest Curtin Economics Centre ('the Centre'). While every effort has been made to ensure the accuracy of this document, the uncertain nature of economic data, forecasting and analysis means that the Centre, Curtin University and/or Bankwest are unable to make any warranties in relation to the information contained herein. Any person who relies on the information contained in this document does so at their own risk. The Centre, Curtin University, Bankwest, and/or their employees and agents disclaim liability for any loss or damage, which may arise as a consequence of any person relying on the information contained in this document. Except where liability under any statute cannot be excluded, the Centre, Curtin University, Bankwest and/or their advisors, employees and officers do not accept any liability (whether under contract, tort or otherwise) for any resulting loss or damage suffered by the reader or by any other person.

The views in this publication are those of the authors and do not represent the views of Curtin University and/or Bankwest or any of their affiliates. This publication is provided as general information only and does not consider anyone's specific objectives, situation or needs. Neither the authors nor the Centre accept any duty of care or liability to anyone regarding this publication or any loss suffered in connection with the use of this publication or any of its content.

#### Authorised Use

© Bankwest Curtin Economics Centre, **November 2019**

Bankwest Curtin Economics Centre Working Paper Series

ISSN: 2202-2791

ISBN: **978-1-925757-04-0**

**Steven Bond-Smith<sup>1</sup>**

<sup>1</sup> **Bankwest Curtin Economics Centre, Curtin University**

#### Suggested Citation

**Steven Bond-Smith, "The unintended consequences of increasing returns to scale in geographical economics." Bankwest Curtin Economics Centre Working Paper 19/04, Perth: Curtin University.**

# The unintended consequences of increasing returns to scale in geographical economics

Steven Bond-Smith

November 2019

## Abstract

Increasing returns to scale is now fundamental to both economics and economic geography. But first generation theories of endogenous growth imply an empirically-refuted scale effect. This scale effect and assumptions to negate the scale effect both imply unintentional spatial consequences. A review of the broad economic geography literature reveals the widespread use and misuse of first generation and semi-endogenous growth techniques despite these distortions. Techniques are suggested for avoiding these unintended spatial consequences. Crucially, the scale-neutral Schumpeterian branch of endogenous growth theory enables research in economic geography to focus on the distinctly spatial mechanisms that define the spatial economy.

## 1 Introduction

Increasing returns to scale is the basis of many powerful results in economics and economic geography. The interplay of increasing and decreasing returns to scale translates into forces for concentration, dispersion and spatial sorting that determine the spatial organization of economies (Proost and Thisse, 2019). Increasing returns to scale is also fundamental to endogenizing growth, because productivity depends only on the stock of non-rival ideas rather than its division between people. But increasing returns to scale in first-generation endogenous growth models results in an empirically-refuted scale effect. While the scale effect has been categorically dismissed from endogenous growth theory (Jones, 1995b), it is still prevalent across the economic geography literature. Yet the scale effect and techniques to negate the scale effect amplify concentration and dispersion forces respectively without a spatial micro-foundation in economic geography. This paper explores the extent of scale effects assumptions about growth in economic geography and proposes that scale-neutral growth models enable research to clearly focus on the distinct geographic mechanisms that determine the internal and external economies that define the spatial economy.

The birth of the so-called “new” growth theory emphasized the role of non-rival knowledge spillovers for growth as an input to and output from endogenous

investment in innovation in return for temporary monopoly profits (Romer, 1990). The spatial nature of knowledge spillovers (Audretsch and Feldman, 1996) and similarities in modelling techniques implied a suitable marriage between endogenous growth and economic geography (Bond-Smith and McCann, 2014). The continued development of endogenous growth theories over four generations (Bond-Smith, 2019) is repeatedly followed by spatial versions to understand the geographic implications of endogenous growth.

Increasing returns to scale is unique in endogenous growth theory because it shows up twice: the scale of production of output *and* the scale of production of ideas. Ideas also multiply because they are both the input and output. This implies an empirically-refuted scale effect (Jones, 1995b), that a larger economy grows faster than a smaller economy and that any rival factor increasing in scale is explosive as it makes use of the same growing pool of ideas. Spatial equilibrium is the balance of increasing and decreasing returns to scale, over various scales. The scale effect unintentionally implies greater innovation productivity for larger regions by amplifying any spatial factor that affects innovation productivity. Second generation semi-endogenous growth models negate the scale effect by assuming diminishing returns to cumulative ideas. There is still increasing returns to scale and the scale effect but it is negated in the long-run by these diminishing returns to ideas. Although this leads to the desired result of constant growth in an aspatial growth model, despite a growing population, in geographic space it unintentionally amplifies the effect of concentrating research effort and dispersing ideas without a spatial foundation for such forces. Schumpeterian growth models without scale effects have the potential for a scale-neutral approach (Bond-Smith et al., 2018; Bond-Smith and McCann, 2019) that enables only the included spatial mechanisms to affect the spatial equilibrium rather than implicit assumptions about growth and scale effects.

Alternatively, the unintentional scale effect is often treated as a convenient mechanism to intentionally facilitate agglomeration economies for innovation in urban growth models, rather than examining the specific causes of agglomeration economies as these models intend. Models in spatial equilibrium specifically provide the opportunity to examine the distinctly spatial factors that result in increasing returns and external economies, rather than rely on implicit assumptions about growth. This misuse of the scale effect therefore sidelines the actual mechanisms that lead to agglomeration economies for innovation and define the spatial equilibrium.

A review of the economic geography literature reveals the wide-spread use and misuse of first generation and semi-endogenous growth techniques despite these distortions. These spatial implications are the unintended consequence of modelling assumptions, rather than true characteristics of the world. The term *unintended consequences* typically refers to Robert Merton's (1936) paper describing that for any policy intervention it is inevitable that some outcomes are accidental. In this article, the proverb applies twice: modelling assumptions can lead to (i) accidental conclusions about spatial-economic phenomena that

imply incorrect policies which also result in (ii) outcomes that are accidental.<sup>1</sup> Therefore this article is of vital importance to improving research quality on spatial economic phenomena and resulting policy implications. By taking careful steps, researchers in economic geography can draw stronger conclusions about the specific mechanisms that drive the spatial equilibrium and develop defensible recommendations for planning or policy.

This article serves four main purposes. Firstly, it highlights a fundamental misunderstanding of increasing returns to scale and the resulting scale effect in endogenous growth theory, such that theorists must be clear about the scales in which increasing returns applies. In a series of toy models, the article explores the spatial consequences of assumptions about growth and scale effects, setting aside standard mechanisms such as pecuniary externalities where the spatial implications are clear. Thirdly it reviews the broad economic geography literature for examples of the use and misuse of the scale effect. Finally, this article aims to reduce the risk that economic geographers or spatial economists inappropriately apply aspatial economic theory to spatial phenomena, by offering techniques to avoid unintended spatial consequences, thereby improving the rigour of research in geographic economics. The paper proceeds as follows. Section 2 outlines the role of assumptions about increasing and decreasing returns to scale in both growth and economic geography. Section 3 develops several toy models to understand the causes and consequences of the scale effect, both in the aspatial endogenous growth model and in a simple two-region model. Section 4 surveys the broad economic geography literature to note examples of research affected by such assumptions about scale effects. Section 5 offers techniques to avoid the unintended spatial consequences of aspatial assumptions about scale effects and Section 6 offers some concluding remarks on a revised research agenda.

## 2 Scale assumptions

Assumptions matter. Economic theory is often criticized that its assumptions are unrealistic (Martin and Sunley, 1998; Romer, 2015). All theoretical models in economics and economic geography use assumptions to avoid the complexity of a truly realistic but intractable theory. Some assumptions are intended to represent true facts about the world, but many are entirely unrealistic and used purely for simplicity. Assumptions eliminate specific variables so the model can focus on what matters. Many assumptions don't matter. Some assumptions eliminate negligible factors that have little impact. Other assumptions ignore factors that are of no interest. And some assumptions are simply "modelling tricks" to ensure the model is tractable or behaves in a way that reflects stylized facts. For all of these types of assumptions, realism is secondary to parsimony.

---

<sup>1</sup>Despite this, I do not agree with Merton's overall conclusion that this implies there should be no intervention. An intervention that is less (or more) successful than expected, but ultimately improves outcomes is still better than no intervention at all. The important factor is the distribution of risk around the estimates of costs and benefits. There may also be significant equity reasons to justify an intervention.

Conclusions drawn from these assumptions are indirect because they need to account for the real world context that is absent in the model. But when assumptions are used to directly draw findings, those assumptions are crucial and should represent a realistic foundation on which to draw strong conclusions. In spatial equilibrium, such a foundation is also spatial.

Assumptions about increasing and decreasing returns to scale are now fundamental to theoretical models in both economics and economic geography. Increasing returns to scale recognizes that production of outputs increases in greater proportion than required inputs. In its simplest form it is a fixed cost of production or scale economies that are internal to the firm, but increasing returns can also create external economies. Increasing returns is one of the oldest problems in economics starting with Adam Smith's division of labor. While economists have long understood that businesses often have a declining marginal cost, perfect competition is incompatible with increasing returns. Avinash Dixit and Joseph Stiglitz's (1977) model of monopolistic competition sparked a revolution in economic theory because prices were no longer pinned down by their marginal cost. Such theories must be clear about the scales in which increasing and decreasing returns apply.

Starting with Romer (1990), theoretical models of endogenous growth recognize that ideas are a purely non-rival input to and output from investing in research in return for temporary monopoly profits. Innovation is endogenized because technological discoveries are a direct result of deliberate decisions to search for them. Non-rivalry implies increasing returns to the scale of the firm's production. Non-rivalry is essential to endogenous growth because it means that productivity depends on the stock of ideas, rather than its division between people. Non-rivalry also enabled increasing returns to the scale of aggregate research because a greater number of researchers can make use of the same set of ideas leading to the empirically-refuted scale effect. Second generation semi-endogenous growth models still incur the scale effect, but assume diminishing returns to cumulative ideas. This implies that ideas become more difficult to find as the simplest ideas are discovered first. There is a scale effect on the level of income due to increasing returns to scale in production and gains from expanding the extent of the market. There is still increasing returns to scale for innovation, but it is reduced by diminishing returns to cumulative ideas. In the balanced growth path the net effect implies constant returns because the scale effect is eventually matched by diminishing returns to cumulative ideas. Schumpeterian models of endogenous growth without scale effects allow innovation to expand in two dimensions: new varieties *and* quality improvements (Young, 1998; Peretto, 1998; Dinopoulos and Thompson, 1998; Howitt, 1999). This changes the scale at which increasing returns to research applies. There is increasing returns to the scale of the firm's research effort, but there are now constant returns to the scale of aggregate innovation as greater research effort is spread across additional varieties, eliminating the scale effect. By clearly understanding the scales at which increasing returns applies, Schumpeterian models enable increasing returns to the scale of the firm or industry while eliminating its impact on innovation in aggregate. While Howitt (1999) also uses

two dimensions for innovation to eliminate the scale effect, it is still assumed that there is diminishing returns to cumulative varieties, as a modelling trick to maintain stylized facts about the impact of R&D incentives. As a result, increasing returns to scale applies to research effort at the scale of the firm and decreasing returns to scale applies to research effort in aggregate. In the balanced growth path these two factors still lead to constant growth because changes in scale spread across additional varieties and increase research effort for established varieties.

Economic geography recognizes that increasing returns to scale also acts over some *spatial scale*. Monopolistic competition and increasing returns therefore also sparked a revolution in both trade and economic geography. Krugman (1979) and Helpman and Krugman (1985) showed how this tool could be embedded into a general equilibrium model of trade such that increasing returns to scale had implications for the location and scale of production. Krugman and Venables (1995) extended the trade model to also include intermediates such that firms have increasing returns to the scale of aggregate production due to vertical linkages. Venables (1999) disaggregated production such that industries divide into clusters. Krugman (1991) extended his trade model to enable the mobility of workers such that increasing returns to scale also had implications for the location of people and firms as utility equalizes across space in the spatial equilibrium or finds a corner solution.

Transport costs are a pecuniary externality related to distance that implies decreasing returns to scale, which eventually limit increasing returns to the scale of production and the size of the market that a firm supplies. Competition also leads to decreasing returns that limit increasing returns to the scale of production. For transport costs increasing returns typically applies at the scale of the firm, but with vertical linkages it applies to the scale of aggregate production. For competition, decreasing returns also applies to the scale of aggregate production. In these models, increasing returns to aggregate scales implies productivity benefits for density or concentration while decreasing returns to aggregate scales implies a congestion cost or dispersion force. Carefully applied to the correct scale, assumptions about increasing and decreasing returns to scale offer powerful insights on the spatial economy.

### 3 Two-region endogenous growth models

Endogenous growth theories can be characterized into generations by the scale effect (Jones, 1999; Bond-Smith, 2019). Each generation is followed by its application to core-periphery or trade models to understand the spatial implications of growth. Using toy models, this section shows the unintended spatial implications of assumptions about scale and growth: (i) in first generation endogenous growth models; (ii) in semi-endogenous growth models; and (iii) in Schumpeterian models without scale effects. Micro-details vary, but this analysis distils the essential elements.

Consider an economy where final output is a composite good  $Y_t = A_t L_{Yt}$

where  $L_{Yt}$  is labor used in production and  $A_t$  represents the stock of ideas. Physical and human capital are left aside in these toy models in order to focus on innovation. Analogous equations apply to two regions (referenced by home and foreign) with variables denoted by a tilde where foreign variables are necessary. New ideas are the result of research effort that builds on the stock of ideas to increase productivity. There is free entry for entrepreneurs, but producers must first develop an idea. The analysis here focuses only on the innovation equations as this is the source of the scale effect. The flow of new ideas is given by the function  $\dot{A}_t = f(L_{At}, A_t)$  where  $L_{At}$  is labor devoted to research effort. Along the balanced growth path, a constant share of labor ( $s < 1$ ) is employed in research such that  $L_{At} = sL_t$  where  $L_t = L_{Yt} + L_{At}$ . The basic model is extended to two regions by applying the ideas production function at the regional level and adjusting the intertemporal knowledge spillover by the location of knowledge using the spillover parameter  $\lambda \leq 1$  such that knowledge is transferred more easily by the close proximity of researchers colocating in the same region. For demonstration purposes in these toy models, it is assumed that transport costs are zero such that the only spatial mechanism included in these toy models is this knowledge spillover parameter. With this limitation these toy models are used to examine the impact of various spatial distributions on the balanced growth, rather than investigating only the steady state distribution, in order to understand the effect of assumptions about scale effects on the spatial forces in the model. Migration is not specified, but the toy models here are flexible enough to accommodate many various alternatives.

### 3.1 First Generation endogenous growth

#### 3.1.1 The aspatial model

In first generation models of endogenous growth (Romer, 1990; Grossman and Helpman, 1991; Aghion and Howitt, 1992), firms invest in research effort that builds on existing non-rival ideas to generate new ideas. The function

$$\dot{A}_t = \gamma L_{At} A_t \quad (1)$$

describes the flow of productivity improvements where  $A_t$  is current technology,  $L_{At}$  is labour devoted to researching new technologies, the dot indicates the flow of new technology over time and  $\gamma > 0$  is a parameter for calibration. Output per capita is proportional to the global stock of knowledge  $y_t = \frac{Y_t}{L_t} = A_t(1 - s)$ . Growth in output per capita is found by taking the time derivative and rearranging,  $g_y = \frac{\dot{y}_t}{y_t} = \gamma L_{At} = \frac{\dot{A}_t}{A_t} = \gamma s L$ . As a result, the growth rate of technology  $g_A = \frac{\dot{A}_t}{A_t}$  and per capita output are proportional to population size:

$$g_{At} = \gamma s L_t. \quad (1a)$$

This is the ‘*scale effect*’, where per capita growth is an increasing function of the scale of the economy  $L_t$ . The model requires that the population is constant to avoid an ever-increasing, explosive growth rate.



### 3.1.2 The regional model

There are many examples of first generation endogenous growth models being extended to two regions (see for example Walz (1997); Martin and Ottaviano (1999, 2001); Baldwin and Forslid (2000a,b); Baldwin et al. (2001); Yamamoto (2003); Baldwin and Martin (2004); Davis (2009)). These models imply that there is a spatial constraint on knowledge spillovers that is the source of agglomeration economies. Consider the innovation function in a two region growth model based on the first generation endogenous growth model above:

$$\dot{A}_t = \gamma s L_t (A_{Wt}) (n + (1 - n) \lambda) \quad (2)$$

where  $L_t$  now represents the local regional workforce,  $\lambda \leq 1$  describes how easily the knowledge stock transfers between places,  $A_{Wt}$  represents global technology and  $n$  represents the share of the knowledge stock developed in the home region  $\frac{A_t}{A_{Wt}} = n$  with an analogous equation describing innovation in the foreign region. Since the final good is a composite good made from traded intermediates, home region output per capita is proportional to the global stock of non-rival knowledge  $y_t = \frac{Y_t}{L_t} = A_{Wt} (1 - s)$  where  $s$  and  $L_t$  now refer to local characteristics. Taking the time derivative and rearranging  $g_y = \frac{\dot{y}_t}{y_t} = \gamma \left( s L_t (n + (1 - n) \lambda) + \tilde{s} \tilde{L}_t (n \lambda + 1 - n) \right) = \frac{\dot{A}_{Wt}}{A_{Wt}} = g_{AW}$ . As a result, the growth rates of the two regions equalise in the steady state to the global technology growth rate,

$$g_{AW} = \frac{\dot{A}_{Wt}}{A_{Wt}} = \gamma \left( s L_t (n + (1 - n) \lambda) + \tilde{s} \tilde{L}_t (n \lambda + 1 - n) \right). \quad (2a)$$

If there are global spillovers, there is be no benefit from locating in the larger region. Setting  $\lambda = 1$  still results in the unintentional scale effect in the Romerian growth model but at the global level.

$$g_{AW} = \frac{\dot{A}_{Wt}}{A_{Wt}} = \gamma \left( s L_t + \tilde{s} \tilde{L}_t \right). \quad (2b)$$

To confirm that the scale effect occurs consider what is required for constant technology growth. In a balanced growth path the growth of knowledge spillovers also matches the constant growth rate of technology. Differentiating with respect to time, setting to zero and rearranging implies that constant technology growth requires no population growth. If the population is growing this implies an ever-increasing growth rate in finite time. The scale effect is typically avoided in these two region models by scaling population relative to global population (i.e. setting  $L_t$  to  $\frac{L_t}{L_t + \tilde{L}_t}$ ) and preventing population growth. If  $s_W$  represents the share of the global population involved in research then the scale effect seems to disappear.

$$g_{AW} = \frac{\dot{A}_{Wt}}{A_{Wt}} = \gamma s_W \quad (2c)$$

Yet the scale effect is still there, it is merely hidden by carefully selected parameters for world population (effectively  $L_t + \tilde{L}_t = 1$ ), population growth ( $g_p = 0$ ) and global spillovers ( $\lambda = 1$ ).

The scale effect re-emerges whenever a spatial mechanism is applied to innovation because it exaggerates the impact of such mechanisms. To examine this closely, reconsider Equation 2a. The growth function is made up of two types of factors: (i) spillover factors

$$(n + (1 - n)\lambda) \quad \text{and} \quad (n\lambda + 1 - n) \quad (3)$$

and (ii) scale factors

$$sL_t \quad \text{and} \quad \tilde{s}\tilde{L}_t. \quad (4)$$

Starting from two equal regions, consider the effect on growth of a shift in the distribution of research effort towards the home region (either by migration in a core-periphery model or by shifting innovation activity in a trade or vertical linkages model) holding all else constant. If growth were only a result of the spatial spillover factors then the distribution of effort, or local scale, would not matter for growth, at least in the short run, because growth would be only affected by the current distribution of technology. But changes in the growth rate are an interaction of both spatial mechanisms in the spillover of knowledge *and* the scale of research. This interaction is not removed by considering only the relative population shares because the relative scale also multiplies the knowledge spillover. As a result, models that use a first generation engine of growth unintentionally imply an additional force for agglomeration economies because concentration of research effort affects growth by amplifying the local knowledge spillover. This factor is in addition to any specific spatial mechanisms that actually create agglomeration economies such as increasing returns to the scale of production, transport costs or the mechanism considered here: spatial externalities on the transfer of knowledge.

## 3.2 Semi-endogenous growth without scale effects

### 3.2.1 The aspatial model

Jones (1995b) showed that growth remained constant despite an increasing population and increasing research effort, refuting the predicted scale effect. To resolve this apparent paradox, Jones (1995a); Kortum (1997) and Segerstrom (1998) developed second-generation theories of endogenous growth without scale effects by diminishing innovation productivity for developing cumulative improvements. In these models, the function

$$\dot{A}_t = \gamma L_{At} A_t^\beta \quad (5)$$

describes how productivity improvements diminish with cumulative discoveries. The parameter  $\beta > 0$  describes how research builds on the cumulative stock of ideas, but  $\beta < 1$  implies that it is increasingly difficult to discover additional new ideas. As above, output per capita is proportional to the global stock of

knowledge  $y_t = \frac{Y_t}{L_t} = A_t(1-s)$  and growth in output per capita is found by taking the time derivative and rearranging,  $g_y = \frac{\dot{y}_t}{y_t} = \gamma L A_t A_t^{\beta-1}$ , which is again equal to the growth rate of technology:

$$g_A = \frac{\dot{A}_t}{A_t} = \gamma s L A_t^{\beta-1}. \quad (5a)$$

In this equation, a larger population implies a scale effect, as above, but the rate of growth is diminished by technological advancement. A balanced growth path implies that technology growth is constant. Differentiating with respect to time finds the constant long run growth rate is

$$g_A = \frac{g_p}{\gamma(1-\beta)} \quad (5b)$$

where  $g_p$  is the population growth rate. That is, in the balanced growth path the scale effect of a growing population is eliminated exactly by diminishing innovation productivity with cumulative ideas to reach a constant growth rate that is proportional to population growth. Research effort has short run impacts on growth, but no impact on the long run growth rate because it is eventually matched by diminishing innovation productivity. While research effort is still an endogenous investment decision in return for temporary monopoly profits, these models attract the label “semi-endogenous” growth because research effort has no long run impact on growth. However, the scale effect on effort is not removed from semi-endogenous models, it is simply equal to the impact of decreasing returns to the scale of cumulative knowledge in the balanced growth path. Furthermore, continued innovation requires population growth without any micro-foundation to explain why.

### 3.2.2 The regional model

The semi-endogenous growth model is extended to two regions by applying the innovation production function in each region with spatial externalities for knowledge spillovers (For example see Minniti and Parello (2011) or Fukuda (2017)). Consider the innovation function in the home region in a two region growth model:

$$\dot{A}_t = \gamma L A_t (A_{Wt} (n + (1-n)\lambda))^{\beta} \quad (6)$$

with an analogous equation for the foreign region. Taking the time derivative of per capita output and rearranging  $g_y = \frac{\dot{y}_t}{y_t} = \gamma s L A_{Wt}^{\beta-1} (n + (1-n)\lambda)^{\beta} + \gamma \tilde{s} \tilde{L}_t A_{Wt}^{\beta-1} (n\lambda + 1 - n)^{\beta} = \frac{\dot{A}_{Wt}}{A_{Wt}} = g_{AW}$ . As a result, the growth rates of the two regions equalise in the steady state to the global technology growth rate,

$$g_{AW} = \frac{\dot{A}_{Wt}}{A_{Wt}} = \gamma \left( s L A_{Wt}^{\beta-1} (n + (1-n)\lambda)^{\beta} + \tilde{s} \tilde{L}_t A_{Wt}^{\beta-1} (n\lambda + 1 - n)^{\beta} \right). \quad (6a)$$

If there are global spillovers, there is be no benefit from locating in the larger region. Setting  $\lambda = 1$  leads to the standard growth rate where growth is proportional to global population and inversely proportional to cumulative knowledge

since  $0 < \beta < 1$ :

$$g_{AW} = \frac{\dot{A}_{Wt}}{A_{Wt}} = \gamma A_{Wt}^{\beta-1} (sL_t + \tilde{s}\tilde{L}_t). \quad (6b)$$

As in the aspatial model population still implies a scale effect at the global level for innovation that is counteracted by an inverse scale effect on cumulative global knowledge. The scale effect is avoided entirely by scaling population relative to global population (i.e. setting  $L_t$  to  $\frac{L_t}{L_t + \tilde{L}_t}$ ). If  $s_W$  represents the share of the global population involved in research then the scale effect seems to disappear

$$g_{AW} = \frac{\dot{A}_{Wt}}{A_{Wt}} = \gamma A_{Wt}^{\beta-1} s_W \quad (6c)$$

but the growth rate is still constantly diminishing due to the remaining effect of decreasing returns to cumulative global knowledge. As in the aspatial model, the balanced growth path implies the constant long run growth rate is proportional to population growth in the regions where research occurs  $\frac{dq}{dt} = \frac{g_p s L_t + \tilde{g}_p \tilde{s} \tilde{L}_t}{s L_t + \tilde{s} \tilde{L}_t} \frac{1}{\gamma^{(1-\beta)}}$ . The scale effect is still present, but in the balanced growth path it is eliminated by the inverse scale effect of decreasing returns to cumulative knowledge.

As with first generation models, the scale effect re-emerges when spatial mechanisms are applied to innovation because it amplifies the impact of spatial factors. Examine the growth rate in Equation 6a. The function is now made up of three types of factors: (i) spillover factors

$$(n + (1 - n)\lambda) \quad \text{and} \quad (n\lambda + 1 - n), \quad (7)$$

(ii) scale factors

$$sL_t \quad \text{and} \quad \tilde{s}\tilde{L}_t \quad (8)$$

and (iii) local idea congestion

$$[A_{Wt} (n + (1 - n)\lambda)]^{\beta-1} \quad \text{and} \quad [A_{Wt} (n\lambda + 1 - n)]^{\beta-1}, \quad (9)$$

While these final two factors are equalised in the balanced growth path in the aspatial growth model, this is not necessarily the case in the two-region model if the spatial distribution is unsteady, even if growth is otherwise in the balanced growth path. That is, semi-endogenous assumptions about growth affect the spatial forces in unsteady states. Specifically, idea congestion applies to the observed global scale of ideas while scale factors apply to the local scale of the region, which are not necessarily in unison. Again consider the effect on growth of a shift in the distribution of research effort towards the home region (either by migration or shifting innovation activity) holding all else constant. If the resulting change in growth were only a result of the spatial knowledge spillover factors then the distribution of labour or accumulation of ideas would not matter for growth. But changes in the growth rate are an interaction of spatial mechanisms in the spillover of knowledge, the scale of research (or its distribution) and (over time) idea congestion in the technology leading region since  $n$  is

a function of  $sL_t$  and  $\tilde{s}\tilde{L}_t$ . That is, the scale of aggregate research effort multiplies the impact of the spillover mechanism and idea congestion diminishes it. In unsteady spatial distributions, these spatial forces are not eliminated even if the model is otherwise in the balanced growth path because their effect on growth also depends on the distribution of research effort. As a result, spatial models of semi-endogenous growth incorporate both a scale effect for innovation effort that implies agglomeration economies and an inverse scale effect for cumulative knowledge that implies diseconomies from knowledge congestion that are only balanced in the steady state spatial distribution with a balanced growth path.

### 3.3 Schumpeterian endogenous growth without scale effects

The empirical results in Jones (1995b) are a clear rejection of scale effects but it doesn't provide evidence for *how* scale effects should be removed. Ultimately, Jones (1999) concludes that this requires further empirical testing. With the wealth of evidence now in favour of the Schumpeterian branch of endogenous growth (Bond-Smith, 2019), spatial economists have extended this approach to regional models.

#### 3.3.1 The aspatial model

'Schumpeterian' models of endogenous growth allow ideas to expand in two dimensions: new varieties and firm-specific quality improvements. Each dimension has a linear relationship to research effort. This allows explosive ideas but avoids explosive growth by sharing research effort across additional varieties. These models recognized that population growth leads to an increase in the variety of products whereas productivity relates to the quality of individual products. In these models equations apply at the firm level. Firm  $i$  production is given by  $Y_{it} = A_{it}L_{Yit}$ . If  $F = \eta L$  represents the number of varieties at time  $t$ ,<sup>2</sup> aggregate production is given by  $Y_t = \left[ \int_0^{F_t} Y_{it}^{\frac{1}{\sigma}} di \right]^\sigma$  where  $\sigma$  describes consumers' love of variety. The function

$$\dot{A}_{it} = \gamma L_{Ait} \bar{A}_{it} \quad (10)$$

describes the flow of quality improving ideas for each individual firm  $i$  where

$$\bar{A}_{it} = \int_0^{F_t} \frac{A_{jt}}{F_t} dj \quad (10a)$$

represents economy-wide productivity and knowledge spillovers to firms. In this model growth is dependent on research effort at the firm or sector level. In the symmetric equilibrium output is  $Y_t = F_t^\sigma \bar{A}_t \bar{L}_{Yt}$  where  $\bar{A}$  and  $\bar{L}$  now

<sup>2</sup>In this toy model,  $F$  is proportional to  $L$ . This assumption is not controversial since Laincz and Peretto (2006) show that this is a natural equilibrium outcome. In any case, it is also not a critical requirement. See Peretto (2018).

represent firm averages such that the functional form of first generation models applies at the firm level. Per capita output reduces to  $y_t = \frac{Y_t}{L_t} = F_t^\sigma \bar{A}_t (1 - s)$ . Differentiating and rearranging  $g_y = \frac{\dot{y}_t}{y_t} = (\sigma - 1) g_p + \frac{\dot{\bar{A}}_t}{\bar{A}_t}$ . Rearranging the firm's quality improvement production function finds that the growth rate of technology is

$$g_{A_t} = \frac{\dot{\bar{A}}_t}{\bar{A}_t} = \gamma \frac{s}{\eta}. \quad (10b)$$

Technology growth is dependent on the share of labor devoted to research rather than the quantum of labor or population size. As a result, increased research effort by firms or sectors increases the growth rate of technology, but increases in the scale of population only increase the number of varieties, thereby eliminating the scale effect from first generation models. Theoretical arguments strongly support the Schumpeterian approach to modelling endogenous growth without scale effects (Peretto, 2018; Bond-Smith, 2019).<sup>3</sup> The weight of empirical research now also finds in favor of Schumpeterian growth without scale effects. For example see Zachariadis (2003); Laincz and Peretto (2006); Ha and Howitt (2007); Ulku (2007); Madsen (2008); Ang and Madsen (2011); Madsen et al. (2010); Greasley et al. (2013).

### 3.3.2 The regional model

In the two region model knowledge spillovers are a weighted average of observable economy-wide productivity (See Davis and Hashimoto (2014))

$$\bar{A}_{it} = (n + (1 - n)\lambda) \int_0^{F_t} \frac{A_{jt}}{F_t} dj, \quad (11)$$

but the function defining the flow of productivity improvements for the average firm is otherwise unchanged:

$$\dot{A}_{it} = \gamma L_{Ait} \bar{A}_{it}. \quad (12)$$

Home region output in equilibrium with globally-symmetric firms is  $Y_t = nF_t^\sigma \bar{A}_{Wt} \bar{L}_{Yt}$  where  $\bar{A}_{Wt} = \int_0^{F_t} nA_{it} + (1 - n)\tilde{A}_{it} di$  represent global productivity averages because final consumption is a global composite and  $\bar{L}_{Yt}$  represents local firm averages. Per capita output in the home region reduces as above  $y_t = \frac{Y_t}{L_t} = F_t^\sigma \bar{A}_{Wt} (1 - s)$ . Differentiating with respect to time and rearranging, the growth of per capita output is  $g_y = \frac{\dot{y}_t}{y_t} = (\sigma - 1) g_p + \frac{\dot{\bar{A}}_{Wt}}{\bar{A}_{Wt}}$ . Rearranging the quality improvement production function for home region firms finds that the growth rate of technology is

$$g_{A_{Wt}} = \frac{\dot{\bar{A}}_{Wt}}{\bar{A}_{Wt}} = \gamma \frac{s}{\eta} n (n + (1 - n)\lambda) + \gamma \frac{\tilde{s}}{\eta} (1 - n) (n\lambda + 1 - n). \quad (13)$$

<sup>3</sup>See also the Appendix to Peretto (2018).

Setting  $\lambda = 1$  leads to the standard growth rate

$$g_{At} = \frac{\dot{A}_{Wt}}{A_{Wt}} = \gamma \frac{s}{\eta} n + \gamma \frac{\tilde{s}}{\eta} (1 - n) \quad (14)$$

which is equal to the aspatial growth rate if the share of labour devoted to research is the same in both regions. Growth is now unaffected by the scale effect because technology production is neutral to scale. Similarly, in this toy model there is no dispersion effect from diminishing innovation productivity. As a result, the assumptions about growth have no effect on the spatial equilibrium. Viewed from the other direction, assumptions about the spatial economy have no impact on growth in the absence of spatial mechanisms (For example when  $\lambda = 1$ ). Instead, the spatial equilibrium is determined precisely by the spatial mechanisms that could be included in the model such as imperfect spillovers ( $\lambda < 1$ ), transport costs, congestion costs, rental costs or any other direct mechanism rather than by assumptions about growth and scale. The function is now made up of only two factors: (i) spillover factors:

$$(n + (1 - n)\lambda) \quad \text{and} \quad (n\lambda + 1 - n), \quad (15)$$

and (ii) share factors:

$$\frac{s}{\eta} n \quad \text{and} \quad \frac{\tilde{s}}{\eta} (1 - n) \quad (16)$$

However, even Schumpeterian models can be susceptible to unintentional consequences in the spatial equilibrium if the growth engine is not scale-neutral. In the above toy model, R&D subsidies have no effect on the research effort of individual firms if additional induced research effort is used only to expand the number of varieties. Dinopoulos and Thompson (1998) argue that this can be avoided if R&D activities are defined to focus solely on quality improvements. Alternatively, to retain the stylized fact that long run growth is influenced by proportional support for research effort, Howitt (1999) adds an assumption that it becomes progressively more difficult to develop additional varieties for larger populations. Specifically, quality improving innovations are subject to constant returns to scale but inventing new varieties is subject to decreasing returns. As a result, economy-wide productivity is not neutral to scale because a growing population implies a slowing rate of innovation. This idea can be shown in the above toy model by modifying the number of varieties such that  $F = \eta L^\beta$  where  $\beta < 1$ . In the aspatial model

$$g_{At} = \frac{\dot{A}_t}{A_t} = \gamma \frac{s}{\eta} L^{1-\beta} \quad (17)$$

such that growth slows as population increases because new varieties are increasingly difficult to find.

In the two region variant (Davis and Hashimoto, 2015)

$$g_{At} = \frac{\dot{A}_{Wt}}{A_{Wt}} = \gamma \frac{s}{\eta} n L^{1-\beta} + \gamma \frac{\tilde{s}}{\eta} (1 - n) \tilde{L}^{1-\beta} \quad (17a)$$

with analogous equations for the foreign region. This implies the spatial implication that productivity improvements diminish as a city grows in size because all the simple ideas have been found first, yet such a conclusion should only apply at the global level or in the limited circumstance where transport costs are so high that two regions cannot trade at all. This distortion in the spatial equilibrium is the result of differing assumptions about quality improving innovations and inventing new varieties that are required to retain specific stylized facts in the aspatial model, rather than a micro-founded spatial mechanism. In this version, an ideas congestion factor reemerges at the regional level.

Davis and Hashimoto (2015) argue that product development costs may be higher in larger markets, but these should be therefore based on the size of the global market, rather than the local region. That is, these are costs that would be faced by any supplier according to the size of the overall global market, irrespective of their location in that global market. Other spatial mechanisms may explain why innovation costs differ between cities, such as land rents for lab space, wage costs for scientists to live in an expensive location or congestion costs for scientists travelling to work, but these spatial models are the appropriate tool to directly model such mechanisms rather than rely on implicit biases in assumptions about scale in the innovation production function. Instead Bond-Smith et al. (2018) and Bond-Smith and McCann (2019) use a scale-neutral innovation production function based on Young (1998) such that the spatial equilibrium is only affected by the intended spatial mechanisms of transport costs, knowledge spillovers and related variety clustering.

Most recently Peretto (2018) represents a fourth generation model of endogenous growth (see (Bond-Smith, 2019)). In this model, quality improvements are explosive, but the number of varieties expands in response to explosive quality improvements. A two region variant has not yet followed, but since the innovation production function is scale-neutral it would not lead to any unintentional consequences for the spatial equilibrium.

## 4 Broader impact of aspatial assumptions

Endogenous growth is now widely explored across various sub-fields of geographical economics. However, spatial models that rely on growth theories with either the unintended scale effect or assumptions to remove the scale effect, derive implications for the spatial equilibria and for regional or trade policy that may be a misinterpretation of the aspatial assumptions about scale effects. City-scale may well be a predictor of growth dynamics, but spatial economists and economic geographers limit their findings if the scale effect is simply assumed to facilitate or amplify the spatial mechanisms that lead to scale economies. Similarly, findings may be misguided if aspatial assumptions to negate the scale effect also unintentionally diminish the spatial mechanisms that define the spatial economy. This section explores assumptions about scale effects and growth in the broad economic geography literature.

Similarities in modelling techniques imply a natural combination of endoge-



nous growth models with the *new economic geography (NEG)* (Bond-Smith and McCann, 2014). This is broadly examined in the toy models above by closing the model with footloose labour. Examples in this literature include Walz (1997); Martin and Ottaviano (1999, 2001); Baldwin and Forslid (2000a); Yamamoto (2003); Baldwin and Martin (2004). In two region NEG models with first generation endogenous growth, the scale effect is not solved by limiting population growth to zero because labour mobility affects the scale of innovation and amplifies the effect of spatial knowledge spillovers in unsteady states. Yet these models usually imply that any benefits from agglomeration economies in the model are solely due to local knowledge spillovers (i.e.  $\lambda < 1$ ) rather than this amplification from the scale effect. In an early survey Nijkamp and Poot (1998) compare spatial implications in neoclassical growth with first generation models finding that models with increasing returns are inherently unstable. This instability is characteristic of many models of spatial equilibrium, but the instability is exacerbated by the scale effect. In two region models with semi-endogenous growth, the scale effect is not removed. Diminishing returns to cumulative ideas implies idea congestion or a dispersion effect that reduces any spatial mechanism in the model. Idea congestion matches the impact of the scale effect only in the steady state distribution in the balanced growth path and otherwise implies this dispersion force. For example, Minniti and Parello (2011) or Fukuda (2017) both conclude that there are benefits for innovation when workers disperse between regions due to assumptions related to semi-endogenous growth without a spatial foundation for such conclusions. In Schumpertian models without scale effects the scale effect is removed by allowing innovation to expand in two dimensions: quality improvement and product variety. As a result, the scale effect is removed by spreading a growing population across additional varieties such that growth can be scale neutral. However, Davis and Hashimoto (2015) use Howitt's (1999) assumption that additional varieties are increasingly difficult to develop, implying a dispersion force without a spatially micro-founded mechanism for such a force. As a result, many two region models of endogenous growth inadvertently draw conclusions about spatial forces that are a result of assumptions about growth and scale. Instead a scale-neutral approach (Bond-Smith et al., 2018; Bond-Smith and McCann, 2019) allows spatial mechanisms to deliberately drive the spatial forces in such models.

These issues are not only limited to models with footloose labour. Given the limiting nature of the scale effect, a number of two region *trade models* restrict migration and make other limiting assumptions that avoid differences in the scale of regions such that regions are equal in all other respects. While some two region trade models leap through these additional hoops (Davis, 1998; Baldwin et al., 2001; Baldwin and Forslid, 2000b, 2010; Minerva and Ottaviano, 2010; Baldwin and Harrigan, 2011; Breinlich et al., 2014), such limiting assumptions are not always distinctly defined or made clear. However, with low transport costs, economies of scale for production and innovation can concentrate activity even in absence of footloose labour. As a result, the implicit scale effect still affects the spatial economy by amplifying any spatial mechanism for knowledge spillovers.

Glaeser (2003) defines the subfield, the *New Economics of Urban and Regional Growth*. This field focuses on empirical urban economics research that estimates the determinants of city growth. The impact of endogenous growth in this predominantly empirical field is in reverse. Much of this research finds that big cities tend to grow at the same rates as many small cities, supporting the Schumpeterian branch of endogenous growth theory without scale effects, even at the city level. This reinforces the theoretical discussions above and an expansion of research into the particular spatial mechanisms that drive relationships between knowledge, human capital and the transmission of ideas, while denying a role for scale to influence growth rates. Instead, scale has a level effect on firm size, city size and productivity at the city level, but no effect on their growth.

Despite this evidence supporting a scale-neutral approach to innovation, *urban economics* models typically combine first generation endogenous growth theories into models of cities (for examples see Duranton (2006, 2007)). First generation models inadvertently predict that innovation (and productivity growth) increases with aggregate city-wide research effort. In doing so, these models assume an agglomeration or clustering benefit for innovation without modelling a mechanism for why such benefits occur. To avoid these consequences from the assumed scale effect, these models cannot examine general population growth and require an additional restrictive assumption that each industry locates in one city alone. Similarly, the scale effect may be explicitly assumed to be the source of external economies associated with an agglomeration or clustering mechanism (as in Black and Henderson (1999)), without actually modelling these urban mechanisms directly. Such models are the perfect vehicle for understanding the urban mechanisms that affect productivity growth and facilitate benefits from clustering or agglomeration economies for innovation but surprisingly offer no such insight. One interesting approach for this literature is whether the urban structure itself offers an alternative explanation and mechanism for removing the scale effect by mitigating increasing returns to scale with city-size related commuting costs (Rossi-Hansberg and Wright, 2007). However, the assumed role of scale effects is of critical importance to these results. The scale effect itself is again effectively an assumed mechanism for agglomeration without modelling the causes of increasing returns to agglomeration. Scale-neutral endogenous growth would be useful in a spatial equilibrium urban model to understand how the mechanisms that create agglomeration economies for innovation are extracted by land rents, congestion and commuting costs in larger cities as a balance of the agglomeration and dispersion forces that such mechanisms create.

*Quantitative spatial economics* as defined by Stephen Redding and Esteban Rossi-Hansberg (2017) offer a series of recent research using continuous space. This allows for a description of space, rather than location in economic models and are an ideal tool for understanding spatial mechanisms that affect growth (Rossi-Hansberg, 2019). While most of these models are static, and avoid modelling growth, a subset of this research led by Klaus Desmet and Esteban Rossi-Hansberg utilize a first generation engine of endogenous growth (See for example Desmet and Rossi-Hansberg (2009, 2010, 2012, 2014); Nagy (Nagy);

Desmet et al. (2018)). The authors acknowledge that the scale effect is a limitation in terms of world population growth, but the scale effect is also an implicit agglomeration factor that amplifies spatial mechanisms, rather than modelling the causes of agglomeration economies directly. Desmet et al. (2018) suggest that one solution is defining the cost of innovation as an increasing function of world population, which acts as a modelling trick so that the model is neutral to global scale. A more advanced scale-neutral innovation production function would complicate calculation of the spatial equilibrium but would be neutral to local scale and enable the spatial mechanisms that cause agglomeration and dispersion forces for innovation to be modelled distinctly.

Lastly, the *empirical regional science* literature uses spatial econometric models to understand spillovers between regions (Lesage and Fischer, 2008). By adding spillovers the interpretation of “scale” is somewhat different. Interpreting these empirical results in the context of the scale effect could lead to misleading conclusions about the role and proximity of neighbours. If models are estimated based on linearising a first generation or semi-endogenous model there is an implicit hypothesis that there is likely to be some additional benefit from the scale of the region including its neighbours. But these regressions rarely include the scale of neighbours such that the benefits of both spillovers and local scale may be overestimated to compensate. Empirical analysis that favours this approach may either leave out or over-emphasise key predictors that are correlated with scale. For example, Parent and LeSage (2012) linearise a semi-endogenous model with spatial externalities to investigate determinants of regional innovation capacity. Such regressions should instead focus on the relative rate of, which is neutral to scale, but recognise that the use of scale as an explanatory variable is only to compensate for missing spatial mechanisms related to agglomeration economies. Local scale may well be an important control variable to compensate for missing confounding variables, but the underlying growth mechanism should be neutral to scale in order to estimate the specific spatial mechanisms that create increasing returns to the scale of local innovation effort.

## 5 Response

The effect of scale on innovation and growth is not disputed. Increasing returns is fundamental and the benefits of internal and external economies appear across the innovation and economic geography literatures. At the same time, economic growth is unrelated to scale (Jones, 1995b). The dispute over scale effects is really about how to reconcile these two stylised facts (Bond-Smith, 2019). Recent models of growth allow innovation to be explosive even if growth is not (Peretto, 2018). But economic geographers and spatial economists should seek to find the spatial mechanisms that create these internal and external economies for innovation rather than rely on the accidental consequences of aspatial assumptions in early endogenous growth models. This section summarises how economic geographers and spatial economists might respond to avoid the unintended con-

sequences of the implicit assumptions about scale and growth.

## 5.1 Avoiding unintended consequences

Models in economic geography are built from a combination of aspatial and spatial assumptions. Some assumptions describe microfounded mechanisms and others are required only for analytical convenience. Unintended spatial consequences can be avoided by modelling aspatial aspects in a manner which is neutral to space such that spatial consequences are only a result of intentional spatial mechanisms. Similarly, empirical work can be interpreted correctly in terms of intentional spatial mechanisms rather than implicit scale assumptions. The spatial or aspatial nature and analytical purpose of assumptions should be clearly defined. Assumptions used for analytical convenience should be examined closely to have no impact on direct conclusions. Assumptions that do not meet these strict requirements should be avoided.

Implementing a scale-neutral approach requires a careful examination of assumptions. Aspatial mechanisms (e.g. growth) must be carefully checked for their unintentional spatial implications. In this way, the conclusions drawn can be explicit about the source of spatial phenomena. Researchers should be explicit about the spatial mechanisms in the model and their micro-foundation. For example, spatial externalities for knowledge spillovers are a mechanism that represents one driver of agglomeration economies for innovation and this spatial pattern is strongly microfounded by the empirical literature (Audretsch and Feldman, 1996). Any microfounded argument to assume a scale effect, or inverse-scale effect should be modelled directly with such a microfounded spatial mechanism and should not be implicit within assumptions about innovation or growth. Spatial economic models are the perfect vehicle to model the real causes of such scale effects, such as agglomeration economies, rather than relying on implicit assumptions about growth. Therefore, assumptions about innovation and growth or any other typically aspatial economic phenomena should be neutral to population size. For example, endogenous growth should be modelled without scale effects. Scale effects should be eliminated by product proliferation (as in the Schumpeterian branch of the literature: Young (1998); Dinopoulos and Thompson (1998); Peretto (1998); Howitt (1999)) but avoid an inverse scale assumption (as in Jones (1995a); Howitt (1999)).

The growth model described by Howitt (1999) includes an assumption of diminishing productivity for innovation in additional varieties to retain stylized facts about growth in response to proportional R&D subsidies. However, in a regional model with spatial externalities for knowledge spillovers, the diminishing innovation productivity assumption would also unintentionally imply a dispersion effect. Modelling Howitt (1999) in space should involve a modified version of the model that is neutral to population such that innovation productivity for additional varieties does not diminish with population. Instead a microfounded spatial mechanism for the cost of commercialisation related to market size that would achieve both an understanding of the spatial economy and the desired stylised facts. As a result, spatial consequences would be clearly a result of such

mechanisms rather than unintentionally caused by assumptions about growth.

Spatial conclusions should be directly connected with their causal spatial mechanisms. A requirement to identify the spatial mechanism(s) highlights any spatial consequence that is otherwise unintentionally caused by implicit assumptions. This provides an opportunity to revise the underlying model when unintended spatial consequences are identified. If such a revision is not possible, models with unintended spatial consequences do not have to be disregarded entirely, but affected conclusions can be treated appropriately as limitations of the particular model used for analytical convenience. This allows a focus on the intentional spatial conclusions and their implications.

This approach is already common, but not always standard with respect to spatial assumptions. For example, assumptions in Krugman (1991) about the proportions of skilled and unskilled workers are not a prediction of equal wages but an analytical tool or modelling trick that allows the model to be tractable. The critical factor is that Paul Krugman (1991) is explicit about the nature of this assumption as a modelling trick.

Many assumptions may have intentional spatial implications. For example, assumptions about increasing returns to the scale of production in Krugman (1991) are clearly intended to show that both trade and the location of economic activity are a result of economies of scale. Such intentional mechanisms must be clearly identified to avoid any misinterpretation.

## 5.2 Empirical research

While empiricists often seek to find the underlying relationships between data without a predefined theory, all empirical research has some intuitive hypothesis about the underlying theoretical mechanisms. Empirical research in spatial economics, urban economics, economic geography and regional science may be misinterpreted if the underlying theoretical models include unintentional assumptions about scale and growth. For example, in order to estimate the contribution of various factors to growth, a log transformation of first generation endogenous growth models implies that city size will be a significant explanatory variable. City size may well be predictive of innovation or growth, but in the original model this is attributed to increasing returns to the scale of research effort leading to the scale effect, rather than the underlying spatial mechanisms that actually create agglomeration economies for innovation. City population may be used as a control variable as a proxy for other unobserved spatial factors that are unrelated to the variable of interest, but it is of limited use for drawing strong conclusions about specific agglomeration mechanisms. Its implication is that bigger is better, rather than understanding why larger cities are able to facilitate a higher level of innovation.

Empirical research can follow a similar approach to theory development to ensure the appropriate mechanisms are captured by the model. By using a growth model that is neutral to scale, spatial results can be attributed to their appropriate spatial mechanisms. By identifying a microfoundation, an appropriate explanatory variable or specification can be selected and results interpreted

correctly. For example, industry specialisation and clustering can appropriately capture intra-industry knowledge spillovers, rather than population scale metrics that might otherwise ignore a role for small specialized cities.

Empirical research also offers significant opportunities to improve both empirical and theoretical research by providing supporting evidence of appropriate theoretical models of growth without spatial distortions. While scale effects are often observed in the economic geography literature due to inherently spatial mechanisms, the identification of such mechanisms is critical to improving understanding of both the spatial economy and economic growth processes. For example, recent research identifies distinct mechanisms such as interaction between researchers (Akcigit et al., 2018) to facilitate innovation and growth. Such mechanisms can also be applied to the spatial economy (McCann, 2007) provided the underlying model is neutral to scale. Perhaps most importantly, undertaking such steps means that credible policy recommendations can be connected to their underlying causes rather than to implicit assumptions about scale and growth.

### 5.3 Re-examining existing research

Conclusions drawn from existing research can be evaluated or qualified in light of the issues raised in this article. In doing so, appropriate steps can be taken to avoid unintended spatial interpretations. There are three proposed approaches: Forward linkages, reverse linkages and a direct approach.

The forward linkages approach starts with the underlying assumptions in the model. Initially, all the underlying assumptions in the model should be clearly identified. Deriving the model from its assumptions allows the spatial mechanics of the model related to population, scale or space to be correctly traced through from this initial audit of assumptions. On this basis, the spatial conclusions of the model can be confirmed or qualified based on the spatial microfoundations of its underlying assumptions.

The reverse linkages approach is identical, but in the opposite direction. Starting with the model's spatial conclusions, the original underlying assumptions and mechanisms that lead to those conclusions can be identified by examining each of the parameters affecting each conclusion. In doing so the appropriate restrictions for each conclusion may become clearer. This approach focuses on the relevance of the conclusions enabling modelling tricks and assumptions to be easily permitted if they do not otherwise affect the implications of final conclusions.

The direct approach focuses specifically on the spatial mechanisms included in the model. This is the most flexible approach, as it permits aspatial assumptions, but denies the spatial implications of distinctly aspatial assumptions. Such spatial mechanisms are typically flow functions, such as trade, migration or knowledge spillovers but could also include scale functions such as increasing returns. In deriving the model, the interaction between spatial mechanisms and otherwise aspatial assumptions should be monitored closely to trace whether spatial forces are altered by aspatial assumptions. To the extent that aspatial

assumptions distort the spatial implications, this defines appropriate restrictions on conclusions of the model.

Overall, spatial conclusions are only as strong as the causative and micro-founded assumptions that underpin them. Spatial implications without such a microfoundation can be appropriately qualified.

## 6 Concluding remarks

Research in economic geography, broadly defined, is limited by assumptions about scale effects and growth. In light of the issues highlighted in this article economic geographers and spatial economists can take note of these limitations and use appropriate techniques in order to focus on the distinctly spatial mechanisms that lead to agglomeration economies for innovation. As a result, researchers can draw stronger conclusions about the specific mechanisms that drive the spatial economy and develop defensible recommendations for planning or policy.

Future research can also overcome such limitations. Theoretical models in economic geography and urban economics are the perfect vehicle for examining the spatial mechanisms driving agglomeration economies and diseconomies for innovation and growth. Yet existing research seems inadequate when the assumed scale effect takes the role of agglomeration or unintentionally amplifies spatial factors. Future research in spatial economic fields should utilize tools from the now broad, expanding and empirically supported Schumpeterian branch of endogenous growth theory, or subsequent generations (Bond-Smith, 2019), where the scale effect is eliminated by product proliferation.

Given the important role of increasing returns to scale for the spatial economy, opportunities for progressing research on endogenous growth in the spatial economy are profound. While existing research can be appropriately qualified, there is ample opportunity to also utilise a more modern understanding of endogenous growth such as Peretto (2018), Akcigit and Kerr (2018) or Akcigit et al. (2018) with appropriate modifications to ensure scale-neutrality. Bond-Smith et al. (2018) and Bond-Smith and McCann (2019) make initial progress by utilising a scale-neutral approach to endogenous growth in two-region models, but limitations in Young (1998) limit their focus to the spatial mechanisms in the model. More complex endogenous growth models that are scale-neutral offer greater opportunity to understand the spatial consequences of innovation directly. There is much progress to be made.

## References

- Aghion, P. and P. Howitt (1992, March). A model of growth through creative destruction. *Econometrica* 60(2), 323–51.
- Akcigit, U., S. Caicedo, E. Miguelez, S. Stantcheva, and V. Sterzi (2018). Danc-

- ing with the Stars: Innovation Through Interactions. NBER Working Papers 24466, National Bureau of Economic Research, Inc.
- Akcigit, U. and W. R. Kerr (2018). Growth through Heterogeneous Innovations. *Journal of Political Economy* 126(4), 1374–1443.
- Ang, J. and J. Madsen (2011). Can second-generation endogenous growth models explain productivity trends and knowledge production in the asian miracle economies. *Review of Economics and Statistics* 93(4), 1360–1373.
- Audretsch, D. B. and M. P. Feldman (1996). R&D Spillovers and the geography of innovation and production. *American Economic Review* 86(3), 630–40.
- Baldwin, R. and J. Harrigan (2011). Zeros, Quality, and Space: Trade Theory and Trade Evidence. *American Economic Journal: Microeconomics* 3(2), 60–88.
- Baldwin, R. E. and R. Forslid (2000a). The core-periphery model and endogenous growth: stabilizing and destabilizing integration. *Economica* 67(267), 307–24.
- Baldwin, R. E. and R. Forslid (2000b). Trade liberalisation and endogenous growth: A q-theory approach. *Journal of International Economics* 50(2), 497–517.
- Baldwin, R. E. and R. Forslid (2010). Trade liberalization with heterogeneous firms. *Review of Development Economics* 14(2), 161–176.
- Baldwin, R. E. and P. Martin (2004). Agglomeration and regional growth. In J. V. Henderson and J. F. Thisse (Eds.), *Handbook of Regional and Urban Economics*, Chapter 60, pp. 2671–2711. Elsevier.
- Baldwin, R. E., P. Martin, and G. I. Ottaviano (2001). Global income divergence, trade and industrialisation: The geography of growth take-offs. *Journal of Economic Growth* 6, 5–37.
- Black, D. and V. Henderson (1999). A theory of urban growth. *Journal of Political Economy* 107(2), 252–284.
- Bond-Smith, S. (2019). The decades-long dispute over scale effects in the theory of economic growth. *Journal of Economic Surveys*, 1–29.
- Bond-Smith, S. and P. McCann (2014). Incorporating space in the theory of endogenous growth: Contributions from the new economic geography. In M. M. Fischer and P. Nijkamp (Eds.), *Handbook of Regional Science*, pp. 213–236. Springer Berlin Heidelberg.
- Bond-Smith, S., P. McCann, and L. Oxley (2018). A regional model of endogenous growth without scale assumptions. *Spatial Economic Analysis* 13(1), 5–35.



- Bond-Smith, S. C. and P. McCann (2019). A multi-sector model of relatedness, growth and industry clustering. *Journal of Economic Geography*. lbz031.
- Breinlich, H., G. I. Ottaviano, and J. R. Temple (2014). Regional Growth and Regional Decline. In *Handbook of Economic Growth*, Volume 2 of *Handbook of Economic Growth*, Chapter 4, pp. 683–779. Elsevier.
- Davis, C. and K.-i. Hashimoto (2014). Patterns of technology, industry concentration, and productivity growth without scale effects. *Journal of Economic Dynamics and Control* 40(C), 266–278.
- Davis, C. and K.-I. Hashimoto (2015). Industry concentration, knowledge diffusion and economic growth without scale effects. *Economica* 82(328), 769–789.
- Davis, C. R. (2009). Interregional knowledge spillovers and occupational choice in a model of free trade and endogenous growth. *Journal of Regional Science* 49, 855–876.
- Davis, D. R. (1998). The Home Market, Trade, and Industrial Structure. *American Economic Review* 88(5), 1264–76.
- Desmet, K., D. K. Nagy, and E. Rossi-Hansberg (2018). The geography of development. *Journal of Political Economy* 126(3), 903–983.
- Desmet, K. and E. Rossi-Hansberg (2009). Spatial growth and industry age. *Journal of Economic Theory* 144(6), 2477–2502.
- Desmet, K. and E. Rossi-Hansberg (2010). On Spatial Dynamics. *Journal of Regional Science* 50(1), 43–63.
- Desmet, K. and E. Rossi-Hansberg (2012). Innovation in space. *American Economic Review* 102(3), 447–52.
- Desmet, K. and E. Rossi-Hansberg (2014). Spatial Development. *American Economic Review* 104(4), 1211–43.
- Dinopoulos, E. and P. Thompson (1998). Schumpeterian Growth without Scale Effects. *Journal of Economic Growth* 3(4), 313–35.
- Dixit, A. K. and J. E. Stiglitz (1977). Monopolistic competition and optimum product diversity. *American Economic Review* 67(3), 297–308.
- Duranton, G. (2006). Some foundations for Zipf’s law: Product proliferation and local spillovers. *Regional Science and Urban Economics* 36(4), 542–563.
- Duranton, G. (2007). Urban evolutions: The fast, the slow, and the still. *American Economic Review* 97(1), 197–221.
- Fukuda, K. (2017). The effects of globalization on regional inequality in a model of semi-endogenous growth and footloose capital. *Asia-Pacific Journal of Accounting & Economics* 24(1-2), 95–105.

- Glaeser, E. L. (2003). The new economics of urban and regional growth. In G. L. Clark, M. S. Gertler, and M. P. Feldman (Eds.), *The Oxford Handbook of Economic Geography*. Oxford University Press.
- Greasley, D., J. B. Madsen, and M. E. Wohar (2013). Long-run growth empirics and new challenges for unified theory. *Applied Economics* 45(28), 3973–3987.
- Grossman, G. M. and E. Helpman (1991). Quality ladders in the theory of growth. *Review of Economic Studies* 58(1), 43–61.
- Ha, J. and P. Howitt (2007). Accounting for trends in productivity and R&D: A Schumpeterian critique of semi-endogenous growth theory. *Journal of Money, Credit and Banking* 39(4), 733–774.
- Helpman, E. and P. Krugman (1985). *Market structure and foreign trade: Increasing returns, imperfect competition, and the international economy*. MIT press.
- Howitt, P. (1999). Steady endogenous growth with population and r & d inputs growing. *Journal of Political Economy* 107(4), 715–730.
- Jones, C. I. (1995a). R&d-based models of economic growth. *Journal of Political Economy* 103(4), 759–84.
- Jones, C. I. (1995b). Time Series Tests of Endogenous Growth Models. *The Quarterly Journal of Economics* 110(2), 495–525.
- Jones, C. I. (1999). Growth: With or Without Scale Effects? *American Economic Review* 89(2), 139–144.
- Kortum, S. S. (1997). Research, Patenting, and Technological Change. *Econometrica* 65(6), 1389–1420.
- Krugman, P. (1979). Increasing returns, monopolistic competition, and international trade. *Journal of International Economics* 9(4), 469–479.
- Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy* 99(3), 483–99.
- Krugman, P. and A. J. Venables (1995). Globalization and the inequality of nations. *The Quarterly Journal of Economics* 110(4), 857–80.
- Laincz, C. and P. Peretto (2006). Scale effects in endogenous growth theory: an error of aggregation not specification. *Journal of Economic Growth* 11(3), 263–288.
- Lesage, J. P. and M. M. Fischer (2008). Spatial Growth Regressions: Model Specification, Estimation and Interpretation. *Spatial Economic Analysis* 3(3), 275–304.

- Madsen, J. (2008). Semi-endogenous versus Schumpeterian growth models: testing the knowledge production function using international data. *Journal of Economic Growth* 13(1), 1–26.
- Madsen, J., J. Ang, and R. Banerjee (2010). Four centuries of British economic growth: the roles of technology and population. *Journal of Economic Growth* 15(4), 263–290.
- Martin, P. and G. I. Ottaviano (1999). Growing locations: Industry location in a model of endogenous growth. *European Economic Review* 43(2), 281 – 302.
- Martin, P. and G. I. P. Ottaviano (2001). Growth and agglomeration. *International Economic Review* 42(4), 947–68.
- Martin, R. and P. Sunley (1998). Slow convergence? the new endogenous growth theory and regional development. *Economic Geography* 74(3), 201–227.
- McCann, P. (2007). Sketching out a model of innovation, face-to-face interaction and economic geography. *Spatial Economic Analysis* 2(2), 117–134.
- Merton, R. K. (1936). The unanticipated consequences of purposive social action. *American Sociological Review* 1(6), 894–904.
- Minerva, G. A. and G. I. Ottaviano (2010). *Handbook of Regional Growth and Development Theories*, Chapter Endogenous growth theories: Agglomeration benefits and transportation costs. Cheltenham, England: Edward Elgar.
- Minniti, A. and C. P. Parello (2011). Trade integration and regional disparity in a model of scale-invariant growth. *Regional Science and Urban Economics* 41(1), 20–31.
- Nagy, D. K. City location and economic development. Unpublished manuscript.
- Nijkamp, P. and J. Poot (1998). Spatial perspectives on new theories of economic growth. *The Annals of Regional Science* 32(1), 7–37.
- Parent, O. and J. P. LeSage (2012). Determinants of knowledge production and their effects on regional economic growth. *Journal of Regional Science* 52(2), 256–284.
- Peretto, P. (1998). Technological change and population growth. *Journal of Economic Growth* 3(4), 283–311.
- Peretto, P. F. (2018). Robust endogenous growth. *European Economic Review* 108, 49 – 77.
- Proost, S. and J.-F. Thisse (2019). What can be learned from spatial economics? *Journal of Economic Literature* 57(3), 575–643.
- Redding, S. J. and E. Rossi-Hansberg (2017). Quantitative spatial economics. *Annual Review of Economics* 9(1), 21–58.

- Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy* 98(5), S71–102.
- Romer, P. M. (2015). Mathiness in the theory of economic growth. *American Economic Review* 105(5), 89–93.
- Rossi-Hansberg, E. (2019). *Geography of Growth and Development*. Oxford University Press.
- Rossi-Hansberg, E. and M. L. J. Wright (2007). Urban structure and growth. *The Review of Economic Studies* 74(2), 597.
- Segerstrom, P. S. (1998). Endogenous Growth without Scale Effects. *American Economic Review* 88(5), 1290–1310.
- Ulku, H. (2007). R&D, innovation, and growth: evidence from four manufacturing sectors in OECD countries. *Oxford Economic Papers* 59(3), 513–535.
- Venables, A. J. (1999). The international division of industries: Clustering and comparative advantage in a multi-industry model. *Scandinavian Journal of Economics* 101(4), 495–513.
- Walz, U. (1997). Growth and deeper regional integration. *Review of International Economics* 5(4), 492–507.
- Yamamoto, K. (2003). Agglomeration and growth with innovation in the intermediate goods sector. *Regional Science and Urban Economics* 33(3), 335–360.
- Young, A. (1998). Growth without scale effects. *Journal of Political Economy* 106(1), 40–63.
- Zachariadis, M. (2003). R&D, innovation, and technological progress: a test of the Schumpeterian framework without scale effects. *Canadian Journal of Economics* 36(3), 566–586.

The Bankwest Curtin Economics Centre is an independent economic and social research organisation located within the Curtin Business School at Curtin University. The Centre was established in 2012 through the generous support from Bankwest (a division of the Commonwealth Bank of Australia), with a core mission to undertake high quality, objective research on the key economic and social issues of relevance to Western Australia.

The Centre's research and engagement activities are designed to influence economic and social policy debates in state and Federal Parliament, regional and national media, and the wider Australian community. Through high quality, evidence-based research and analysis, our research outcomes inform policy makers and commentators of the economic challenges to achieving sustainable and equitable growth and prosperity both in Western Australia and nationally.

The Centre capitalises on Curtin University's reputation for excellence in economic modelling, forecasting, public policy research, trade and industrial economics and spatial sciences. Centre researchers have specific expertise in economic forecasting, quantitative modelling, microdata analysis and economic and social policy evaluation.

A suite of tailored and national economic models and methods are maintained within the Centre to facilitate advanced economic policy analysis: these include macroeconomic and time series models, micro(simulation) models, computable general equilibrium (CGE) models, spatial modelling methods, economic index analysis, and behavioural modelling methods.

## CONTACT

Bankwest Curtin Economics Centre  
Curtin University  
Kent Street Bentley WA 6102  
GPO Box U1987 Perth WA 6845

Tel: +61 8 9266 2873

[bankwesteconomicscentre@curtin.edu.au](mailto:bankwesteconomicscentre@curtin.edu.au)  
[business.curtin.edu.au/bcec](http://business.curtin.edu.au/bcec)