

# Do regional economists answer the right questions?

## On the current discrepancy between the questions regional economists solve and the questions policy makers actually ask

Thomas de Graaff<sup>1\*</sup>

### Abstract

This position paper revolves around two main propositions: namely, (i) regional (or spatial) economists are very restrictive in the tool set they apply, and consequently (ii) their models do not always match with the type of questions policy makers are concerned about. To start with the latter, policy makers—whether national, regional or local—are oftentimes concerned about holistic approaches and future predictions. Exemplary questions are “What works best for my country/region/city” and “If we change this what will happen to the country/region/city as a whole?”. Regional economists—actually, most economists—usually isolate phenomena in order to, at best, explain the impact of a single determinant. Indeed, most regional economists feel very uncomfortable when asked to predict or give the best set of determinants for a certain phenomenon. This has its consequences for the tool set regional economists apply. Usually a parametric regression type of framework is applied isolating the determinant under consideration and controlling as much as possible for observables and unobservables, ideally in a pseudo-experimental framework. A direct consequence of this approach is that emphasis is very much on explaining the impact of an isolated determinant and not on predicting (non-marginal) changes in larger systems. For many applications that is definitely the right approach. However, as this paper ultimately argues, it is very much as well a selective approach that does not do well to deliver on some of the questions policy makers ask regional economists.

### Keywords

Regional science — spatial heterogeneity — conditional robustness — predicting — data science

<sup>1</sup> Department of Spatial Economics, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

\*Corresponding author: ✉ t.de.graaff@vu.nl; 📧 thomasdegraaff.nl

## Contents

<b>Introduction: two different cultures</b>	<b>1</b>
<b>1 The road to current day's practice</b>	<b>2</b>
<b>2 Regional economists turning the blind eye</b>	<b>2</b>
2.1 The blind eye in research . . . . .	2
2.2 The blind eye in education . . . . .	3
<b>3 Incorporating the data science culture</b>	<b>3</b>
3.1 For research . . . . .	3
Regional heterogeneity • Conditional robustness • Regional sorting models	
3.2 For education . . . . .	3
<b>4 Into the abyss</b>	<b>3</b>
<b>References</b>	<b>3</b>

## Introduction: two different cultures

The sexiest job in the next 10 years will be statisticians.

Hal Varian, 2009

The quote above from Hal Varian is in one aspect wrong;

nowadays, we do not call them statisticians but data scientists instead. Nevertheless, in the last two decades companies such as Google, Ebay, Whatsapp, Facebook, Booking.com and Airbnb, have not only witnessed enormous growth but to a large extent also changed the socio-economic landscape. Indeed, with the increasing abundance of (spatial) data and computer capacity, the ability to gather, process, and visualize data has become highly important and therefore highly in demand as well. And all the models and tools these data scientists within these companies use are very much *data driven* with often remarkable results.

In his controversial and path-breaking article, Breiman (2001) presented two different cultures in statistical science. One governed by a (probability) theory-driven modeling approach and one governed by a more (algorithmic) data-driven approach. These two cultures carry over to the econometric and ultimately the empirical regional economics domain<sup>1</sup> as

<sup>1</sup>I use a wide definition for the regional economics domain, which consists of most aspects of regional science in general but for which the theoretical approach is always from an economic perspective. Topics such as, e.g. interregional migration, trade, transport flows and commuting on the one side and regional performance, regional clustering, population growth and specialisation on the other side fall all under this, admittedly, rather wide

well, where—commonly for all social sciences—the theory driven approach still very much dominates the landscape of the realm of contemporary regional economics.

Figure 1 is an adaptation from the one displayed in Breiman (2001) and describes the processes governing these two cultures. Figure (1a) is what I refer to as the modeling approach, where a statistical model is postulated and is central to this culture. This is the classical approach<sup>2</sup> where statistical probability theory meets the empiricism of Karl Popper. Usually the model assumed is stated as a linear model and in its most simple form can be denoted as:

$$y = \mathbf{x}\beta + \varepsilon, \quad (1)$$

where in (regional) economics language,  $\mathbf{x}$  is referred to as the independent variable,  $y$  as the dependent variable and  $\varepsilon$  as a residual term. In this setup, using the data at hand, one typically constructs a statistical test to which extent the estimated coefficient (denoted with  $\hat{\beta}$ ) deviates from a hypothesized value of the coefficient (denoted with  $\beta_0$ )—typically the hypothesis  $H_0 : \hat{\beta} = 0$  is used with as alternative hypothesis that  $H_1 : \hat{\beta} \neq 0$ . However, that is always within the context of the *postulated* model. So, when the null-hypothesis is rejected, it not necessarily means that the true  $\beta$  is unequal to zero, it might also be caused by errors in measuring  $\mathbf{x}$  or even using the wrong *model*!<sup>3,4</sup>

Therefore it is striking that in economics in general, and in regional economics in specific, most of the tools employed are very much *theory or model driven* instead of data driven. My (conservative) estimate would be that 90% of all empirical work in regional economics revolves around postulating a (linear) model and testing whether (a) key determinant(s) are significantly different from a hypothesized value—usually zero.<sup>5</sup> That is, *within* the context of the model assumed.

At best, this approach can be seen in a causal inference framework. If a determinant (such as a policy in the context of economics)  $x$  changes, does it cause then a change in the output  $y$  (most economists typically use some welfare measure).<sup>6</sup>

umbrella.

<sup>2</sup>Sometimes as well referred to as the frequentists' approach. However, this typically concerns the debate between classical statistics and Bayesian statistics, where the two approaches I refer to are more concerned with wider frameworks, of which the Bayesian approach is just one of the elements.

<sup>3</sup>One of the assumptions for regression techniques such as the one used here is actually no misspecification of the model, but—apart from some possible tests on the functional form *within* a specific regression form—usually little attention is given on the validity of the model used. More importantly, within this framework the model itself is usually not tested *a posteriori*.

<sup>4</sup>There is another fallacy with this approach that is often overlooked and that is that the alternative hypothesis being true is a probability as well. Namely, most hypotheses researchers test are typically not very probable. Not taken this into account would actually lead to more null hypotheses to be rejected than should be (false positives).

<sup>5</sup>In a seminal contribution, Breiman, 2001 states that deep into the 90s 98% of the statisticians actually employed the theory driven paradigm and only 2% a data driven paradigm. With the advent of the availability of internet connectivity, large (online) data sources, and faster computers the statistical realm changed dramatically. However, this has not permeated yet in the social sciences.

<sup>6</sup>Most of this research actually intends to mimic a *difference-in-difference*

This approach thus provides a rigid and useful approach to regional policy evaluation. If we implement policy  $x$ , does welfare measure  $y$  then improve?

However, policy makers oftentimes have different questions for which they need solutions. Usually, they revolve around questions starting with “*What determines performance measure A?*” or, more generally, “*What works for my region?*”. These types of questions require a different approach than the previous one. Namely, the former type requires an approach focused on **explaining** while the latter type requires an approach focused on **predicting**.

Figure 1b)

The remaining part of this position paper is structured as follows. The next section deals with the historical background both from an applied statistical and econometric point of view and from a regional science point of view. Section 2 deals with current practices and describes the ‘traditional’ inference based approach as well as some data-driven approaches that have been used in the recent past (though by far not as often as the traditional methods). Section 3 sets out both a research and an education agenda as it addresses how to bridge the gap between the daily practices of regional economists and the demands of local policy makers. The final section shortly summarizes the main points raised in this position paper.

## 1. The road to current day's practice

Regional science itself started with the seminal work of

This early work was very holistic in approach—although mostly from a modeling perspective. Also, regional science was not so much a field on its own dominated by economic and socio-geographic paradigms but very much entwined with statistics, transportation science, sociology and other related fields.

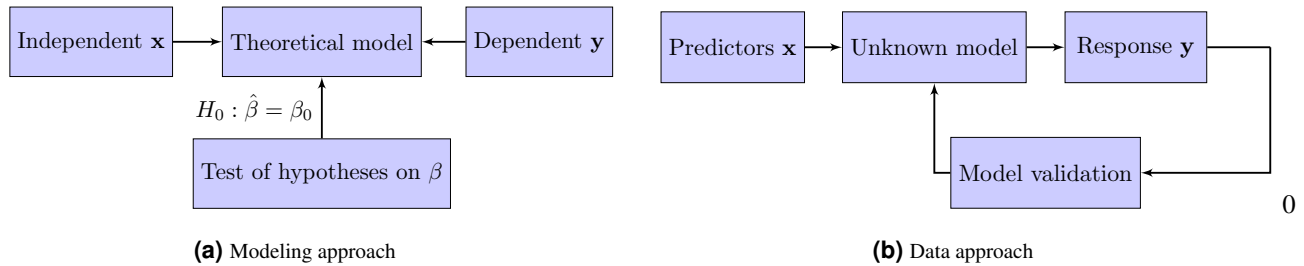
## 2. Regional economists turning the blind eye

### 2.1 The blind eye in research

Regions are conceptually different than cities, as they contain urban, suburban and rural areas simultaneously. Whilst smaller regions can still be seen as the total influence radius of metropolitan areas—such as measured by the concept of local labor markets and the NUTS-2 regions in Europe—, larger regions can typically contain a multiple of cities in combinations with their various hinterlands—such as the Dutch Randstad, the Belgian Flemish Diamond, and the German Ruhr areas.

In recent years, the urban economics literature witnessed large growth; not only noticed by a wider acceptance in mainstream economics, but as well by larger scientific rigor and increased robustness of empirical findings. Remarkably, the empirical regional economics (or regional science in general for that matter) literature lagged behind, although many

approach and gained enormous momentum with the textbook of Angrist and Pischke (2008).



**Figure 1.** Two cultures of statistical/econometric modeling (inspired by Breiman, 2001)

concepts and challenges in both disciplines are conceptually similar and are derived from similar theoretical backgrounds.

Similar to urban economics, there is not yet a clear (consensus in) understanding in which policy instruments are actually (cost-)effective in promoting regional growth.

To do so, I first review the previous literature in section 1. This section focuses mainly on regional economics as it has a larger emphasis on *causal* effects. To a lesser extent we deal with the (economic geography) literature. Based on this literature review Section ?? deals with the research gaps that can be identified.

- housing & population;
- amenities;
- connectivity & accessibility
- networks;
- social & human capital

## 2.2 The blind eye in education

### 3. Incorporating the data science culture

What do we need?

#### 3.1 For research

##### 3.1.1 Regional heterogeneity

(Thissen et al., 2016; de Graaff et al., 2012b,a)

##### 3.1.2 Conditional robustness

In regional science in general and in regional economics in specific, remarkably little attention has been given to reproducibility and robustness of results (with some exceptions as, amongst some others, by Rey, 2014; Arribas-Bel and de Graaff, 2015; Arribas-Bel et al., Forthcoming).

##### 3.1.3 Regional sorting models

As in Bayer et al. (2004) and Bayer and Timmins (2007) and recently by Zhiling et al. (2016) and Bernasco et al. (Forthcoming).

#### 3.2 For education

Schwabish (2014)

## 4. Into the abyss

### References

- Angrist, J. D. and J.-S. Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.
- Arribas-Bel, D. and T. de Graaff (2015). “WooW-II: Workshop on open workflows”. In: *REGION 2.2*, pp. 1–2.
- Arribas-Bel, D., T. de Graaff, and S. Rey (Forthcoming). “Looking at John Snow’s cholera map from the XXIst Century: a practical primer on reproducibility and Open Science”. In: *Regional Research Frontiers: The Next 50 Years*. Ed. by R. Jackson and P. Schaeffer. Berlin: Springer.
- Bayer, P. and C. Timmins (2007). “Estimating Equilibrium Models Of Sorting Across Locations”. In: *The Economic Journal* 117.518, pp. 353–374.
- Bayer, P., R. McMillan, and K. Rueben (2004). “An Equilibrium Model of Sorting in an Urban Housing Market”. NBER working paper: No. w10865.
- Bernasco, W., T. de Graaff, J. Rouwendal, and W. Steenbeek (Forthcoming). “Social Interactions and Crime Revisited: An Investigation Using Individual Offender Rates”. In: *Review of Economics and Statistics*.
- Breiman, L. (2001). “Statistical Modeling: The Two Cultures”. In: *Statistical Science* 16.3, pp. 199–231.
- De Graaff, T., F. G. van Oort, and R. J. Florax (2012a). “Regional Population-Employment Dynamics Across Different Sectors of the Economy”. In: *Journal of Regional Science* 52.1, pp. 60–84.
- (2012b). “Sectoral heterogeneity, accessibility and population-employment dynamics in Dutch cities”. In: *Journal of Transport Geography* 25, pp. 115–127.
- Rey, S. J. (2014). “Open regional science”. In: *Annals of Regional Science* 52.3, pp. 825–837.
- Schwabish, J. A. (2014). “An economist’s guide to visualizing data”. In: *The Journal of Economic Perspectives* 28.1, pp. 209–233.
- Thissen, M., T. de Graaff, and F. G. van Oort (2016). “Competitive network positions in trade and structural economic growth: A geographically weighted regression analysis

for European regions”. In: *Papers in Regional Science* 95.1, pp. 159–180.

Zhiling, W., T. de Graaff, and P. Nijkamp (2016). “Cultural Diversity and Cultural Distance as Choice Determinants of Migration Destination”. In: *Spatial Economic Analysis* 11.2, pp. 176–200.