Multiple models, one explanation

Chiara Lisciandra & Johannes Korbmacher

Published online: 05 Mar 2021.

Submit your article to this journal

View related articles

View Crossmark data

Routledge
Taylor & Francis Group

Multiple models, one explanation

Chiara Lisciandra[a] and Johannes Korbmacher[b]

[a]Faculty of Economics and Business, University of Groningen, Groningen, Netherlands; [b]Department of Philosophy and Religious Studies, Utrecht University, Utrecht, Netherlands

ABSTRACT

We develop an account of how mutually inconsistent models of the same target system can provide coherent information about the system. Our account makes use of ideas from the debate surrounding robustness analysis and draws on the idea of a *shared structure* among models. To illustrate, we consider a case study from international trade-theory.

1. Introduction

In scientific practice, we often find situations where scientists use two or more mutually inconsistent models together to study one and the same target system. Morrison (2011), for example, points out that in physics, we can currently find over 30 mutually inconsistent models of the atomic nucleus in use. Parker (2010a, 2010b), to give another example, discusses the use of so-called *ensemble prediction methods* in climate science, where varieties of model structures under different initial conditions and parameter settings are used to make weather or climate predictions about the same climate system.

This use of families of mutually inconsistent models presents us with an epistemological problem, known in the literature as *the problem of inconsistent models*, see Chakravartty (2010), Massimi (2018), and Morrison (2011, 2015):[1]

1 *The Problem of Inconsistent Models.* How can several mutually inconsistent models together provide coherent information (explanations, predictions, …) about their shared target system?

The aim of this paper is to formulate a novel, broadly *structural* approach to the problem of inconsistent models. The main advantage of our approach is that it can even be applied in situations where models of one and the same phenomenon differ in their most fundamental assumptions.

Existing accounts of how families of mutually inconsistent models can provide coherent information typically address the problems arising from inconsistency by relativizing the use of models to *contexts*. The idea is that if each model in an inconsistent family is used in different kinds of contexts, for different purposes, then the mutual inconsistency of the models no longer poses an epistemological problem. That is, for example, the approach of perspectival realists, like Giere (2006), Chakravartty (2010), and Massimi (2018); or of model pluralists, like Aydinonat (2018) and Grüne-Yanoff and Marchionni (2018).

But such approaches reach their limits when different mutually inconsistent models are used to account for *one and the same phenomenon* concerning their shared target system. Think, for

---

instance, of Parker's analysis of ensembles of climate models being used to predict the weather of a given climate system. Here, there don't seem to be different contexts which allow us to avoid the problem arising from the inconsistencies between the models in the ensemble.

Parker (2010a, 2010b) argues that in the case of ensemble prediction methods, the point is to show which results are invariant across a set of mutually inconsistent models, in order to reduce *uncertainty* that derives from not knowing the exact model structure, precise initial conditions or specific parameter settings. But note that on an approach like Parker's, all we get is information about the *probability* of a certain outcome – in this case, climate predictions. In contrast, the approach that we develop in this paper allows us to extract *explanatory* information from classes of mutually inconsistent models: a shared explanation of a single phenomenon across potentially fundamentally different models.

In order to formulate our approach, we will draw inspiration from the philosophical debate on *robustness analysis*, i.e. the scientific method of (ostensibly) confirming results by showing that they are invariant across different, possibly mutually inconsistent models (Lisciandra, 2017). Weisberg (2006, 2012) proposes an account of robustness analysis, according to which we establish so-called *robust theorems* of the form:

- '*Ceteris paribus*, if [common causal structure] obtains, then [robust property] will obtain.' (Weisberg, 2006, p. 738)

The idea that we will pursue in this paper is that if we can establish a robust theorem with respect to a class of models, then we can take the claim that the common structure explains the robust property to be shared information across these models. In other words, while the models may disagree on all sorts of things, they agree on the claim that the shared structure explains the robust property.

While we draw on ideas from robustness analysis, we'd like to emphasize that we're primarily interested in *in*formation not *con*firmation. That is, we're trying to provide an account of how mutually inconsistent models can coherently *say* something about a shared target system, not an account of whether what they say is true, confirmed, or the like. One might want to argue in the spirit of Ylikoski and Kuorikoski (2010), for example, that if an explanation is shared across fundamentally different models, this shows that the explanation is non-sensitive and thus has greater explanatory power. Indeed, following Lehtinen (2016, 2018) one might argue that the robust result receives (indirect) confirmation from its robustness. Such claims, however attractive, are *orthogonal* to our purposes in this paper: though the claims are compatible with what we're saying, they go in another direction.

In order to illustrate our approach to inconsistent models, we will consider a case study from international trade theory: the case of the so-called *gravity equations* (Anderson, 2011; Head & Mayer, 2014). These equations model international trade flows roughly analogously to Newtonian gravity: the trade flow between two countries is proportional to the product of their economic 'masses' (in the simplest case: GDP) and inversely proportional to their (square) distance. Gravity has been introduced into trade theory as an econometric model by Tinbergen (1962). Anderson describes it as 'one of the most successful empirical models in economics' (Anderson, 2011, p. 133). Theoretical foundations seemed shaky at first – the first theoretical justification for gravity was given by Anderson (1979) only in 1979 – but today derivations in various competing, mutually inconsistent trade models are known, see, e.g. Eaton and Kortum (2002), Bergstrand (1985), Deardorff (1998), Anderson (1979), and Anderson and Van Wincoop (2003)[2]

In our case study, we'll take a closer look at the so-called *structural* gravity equations, a special form of the gravity equations introduced by Anderson and Van Wincoop (2003).[3] What makes structural gravity a fruitful case-study for our approach is that structurally identical, *isomorphic* gravity equations can be derived in many different, mutually inconsistent trade models: despite making fundamentally different assumptions, these models share a common 'gravity structure.' We'll look at how this gravity structure is used by Anderson and Van Wincoop (2003), Arkolakis et al. (2012),

and Allen et al. (2020) to provide explanatory results. These explanations, we argue, are examples of shared explanatory information across the different, mutually inconsistent models.

What sets these structural gravity explanations apart from other robust theorems in economics, such as the ones discussed in Kuorikoski et al. (2010), is that the gravity structure used in the explanations is not the core causal structure of the models involved. The models are built around different trade-theories and mechanisms explaining international trade flows. But, as it turns out, they all instantiate one and the same gravity structure, albeit in different ways.

Here is an overview of the rest of the paper. In Section 2, we shall briefly clarify our view on scientific models and what it means for two models to be mutually inconsistent. In Section 3, we outline the specific version of the problem of inconsistent models we shall address in this paper. Section 4 contains our case study of the gravity equations in international trade theory. In Section 5, we draw some philosophical conclusions and give an outlook.

2. Models and inconsistency

Following Frigg and Hartmann (2018, Section 1), we can distinguish two fundamentally different uses of scientific models.[4] One function of models is to interpret scientific theories by providing a model of their laws and assumptions (Frigg & Hartmann, 2018, Section 1.3) This use of models is central to the so-called *semantic view* of scientific theories, which roughly states that a scientific theory is a collection of models, e.g. Landry (2007), Suppe (1989), Suppes (1960, 2002), and Van Fraassen (1980). The other function of models is to study selected aspects of the world. A model used with this function has a *target system*: a phenomenon, an object, a complex system of objects, a set of data, or the like.[5] Models of the atom, for example, have the atom as their target system, climate models have a region's climate as their target system, and so on. In this paper, we're primarily concerned with this second use of models, which Frigg and Hartmann call 'representational' (Frigg & Hartmann, 2018, Section 1.1–2).

In the literature, we find several different versions of the representational view as well as alternative accounts. Criteria for representational adequacy include, among others, isomorphism (Van Fraassen, 1980), similarity (Giere, 1988; Teller, 2001; Weisberg, 2012) or partial similarity (Mäki, 2006, 2009). Accounts that are not based on representational functions of models are for instance Morgan and Morrison's (Morgan & Morrison, 1999), who view models as mediators. Knuuttila (2005, 2011) as epistemic artifacts and Grüne-Yanoff as heuristic devices.[6] In this paper, we wish to leave open from what the epistemic value of scientific models derives, be it representational accuracy, explanatory power, or yet something different. For most of the points we wish to raise in this paper, the precise answer to the question doesn't matter and we shall, accordingly, leave it open.

As Frigg and Hartmann point out (Frigg & Hartmann, 2018, Section 1.1), the two functions of models – models of theories and models of target systems – do not exclude each other: a model can be both an interpretation of a scientific theory and function as a scientific device for the study of a target system. Take a simple pendulum model, for example. This a model interprets the laws of mechanics over a simple set-up of a point-mass swinging on a massless cord in some background space-time, and as such, it can be viewed as a model of the theory. But at the same time, the simple pendulum model can also be used to predict the behavior of a real-world pendulum, for example. But the use of scientific models as epistemic devices is broader. As Morrison (1999, pp. 48–53) points out, 'theoretical' models, like the simple pendulum model we just sketched, are typically in need of corrections to become a more realistic model of its target system. These corrections, however, are not driven by considerations coming from the interpreted theory, but rather from the 'real world' behavior of the target system. More generally, Morrison argues in Morrison (1999) that models as scientific devices are *autonomous agents*: they are neither (entirely) derived from nor constitutive of scientific theories.[7]

In scientific practice, models as scientific devices come in many shapes and forms: they come as set-theoretic structures (e.g. space-time structures in physics), as mathematical equations (e.g. general equilibrium models in economics), or even computational algorithms (e.g. evolutionary

models in computational biology). In the following, we shall take this *heterogeneity of models*at face value, i.e. we will reason about models as they are presented in scientific practice. Borrowing a phrase from French (who works on the semantic approach to theories), we 'keep quiet on the ontology of models' (French, 2010).[8]

Let's now turn to the central concept of two models being mutually inconsistent. We can distinguish different ways in which two models of the same target system can be mutually inconsistent depending on where the inconsistency is located.

Perhaps the most direct way concerns models which share essentially the same assumptions, but which have different initial conditions, variable settings, parameter settings, or the like. Think, for example, of a climate simulation run on a range of different initial conditions, each time producing a different result (discussed by Parker, 2010b, p. 988 under the label 'initial condition ensembles'). These models contradict each other with respect to the values of the variables in the initial variable setting and all the consequences of those. We shall call models which are mutually inconsistent in this sense *factually* inconsistent with each other since they agree on the structure of the target system, as it were, while disagreeing about the facts.

More involved ways for different models of the same target system to be mutually inconsistent arise when two models use different assumptions in a more robust sense. But first note that the fact that two models use different representations of the same target system expressed in different assumptions, this doesn't necessarily mean that the models are mutually inconsistent: different models can represent the same target system in different but fundamentally compatible ways (Frigg & Hartmann, 2018, Section 1). Morrison (2011, pp. 344–346), for example, discusses models from fluid dynamics, which make different but fundamentally compatible assumptions about the nature of turbulent fluids, see Morrison (2011, p. 346).[9]

Among models with different incompatible assumptions, we may distinguish relatively harmless cases of mutual inconsistency (at least from our current perspective), viz. cases where idealizing assumptions, tractability assumptions, or the like are concerned. Scientific models typically involve a certain degree of idealization and assumptions made for mathematical tractability, cf. Frigg and Hartmann (2018, Section 1) and Kuorikoski et al. (2010, pp. 547–548). The effect of these kind of auxiliary assumptions on scientific results is studied extensively in the literature on robustness analysis, cf., e.g. Kuorikoski et al. (2010), Lehtinen (2016, 2018), Lisciandra (2017), and Weisberg (2006). But from the perspective of the multiple models problem, the issues raised by conflicting auxiliary assumptions are not particularly pressing. As long as what Kuorikoski et al. (2010, p. 547) call the 'substantial' assumptions of two models are compatible with each other, these substantial assumptions can ground shared information provided by the models. The substantial assumptions of a model describe its 'core causal mechanism' (see Kuorikoski et al., 2010, p. 547), so even if the auxiliaries of two models conflict but their substantial assumptions are compatible, then we can take them to coherently describe the same causal structure. In fact, this idea underlies standard robustness analysis methods, where we test whether a result remains invariant under change of auxiliary assumptions, while keeping the substantial assumptions the same (Kuorikoski et al., 2010; Weisberg, 2006).

If, however, two models of the same target system are based on conflicting substantial assumptions, we're dealing with a kind of inconsistency that directly gives rise to the problem of inconsistent models. If two models are based on different, incompatible substantial assumptions, they ascribe to their target system a different core causal mechanism. This makes it *prima facie* unclear how they can coherently agree on claims that they make about their target system. In the following, we shall call models that are based on different substantial assumptions (in the sense of Kuorikoski et al., 2010), *conceptually* inconsistent models since they appear to conceptualize their target system in fundamentally different, incompatible ways. The models of international trade in our case study, Section 4, are mutually inconsistent in precisely this, conceptual way.

Having discussed the different ways in which pairs of models can be jointly inconsistent, we will now move to a discussion of where standard approaches to the problem of multiple models reach their limits.

3. The problem of multiple models

We'd like to begin by pointing out that the problem of inconsistent models is particularly pressing if we take a realist view of models, according to which the epistemic value of a model derives from its representational accuracy (Frigg & Hartmann, 2018, Section 5.1). But two mutually inconsistent models cannot both accurately represent the same target system. So, on a realist view of models, it's *prima facie* unclear how families of inconsistent models can have any combined epistemic value *at all*.

But there's also a problem if we take a more anti-realist view of models, like the one of Cartwright (1983), according to which the epistemic value of a model derives from its explanatory power. In mainstream epistemology, consistency is standardly assumed to be a necessary condition for coherence.[10] So, there is the obvious question of how a family of mutually inconsistent models can provide a *coherent* explanation concerning their shared target system.

A simple solution to the problem of inconsistent models would be to chalk it up to *scientific disagreement*. The idea would be to say that inconsistent models are used because scientists disagree about the relevant facts (in the case of factual inconsistency) or the structure of the nature of the target system (in the case of conceptual inconsistency). We don't doubt that there are cases where this *is* the correct explanation for why inconsistent models are used. But, at the same time, it's clear that the explanation doesn't work in *all* cases. There are several known examples, both historical and contemporary, where mutually inconsistent models are used by a scientific community which is fully aware of the inconsistencies but continues to use the models side-by-side regardless.

For a historic example, consider the following (widely publicized) quote by Infeld and Einstein on the particle-wave duality in the physical study of light:

> But what is light really? Is it a wave or a shower of photons? There seems no likelihood for forming a consistent description of the phenomena of light by a choice of only one of the two languages. It seems as though we must use sometimes the one theory and sometimes the other, while at times we may use either. We are faced with a new kind of difficulty. We have two contradictory pictures of reality; separately neither of them fully explains the phenomena of light, but together they do. (Einstein & Infeld, 1938, pp. 262–263)

The point here is that (at least relative to the state of physics in 1938) we seem to need two mutually inconsistent models to provide the full scientific picture of a certain phenomenon – light. A similar situation seems to apply in the case of atomic models discussed by Morrison in Morrison (2011) (mentioned in the introduction). And yet another example is the case of ensemble prediction methods discussed by Parker (2010a, 2010b), where inconsistent models are used side-by-side to reduce uncertainty about several aspects of a highly-complex phenomenon. Also in economics, as Aydinonat (2018) points out, the situation where economists work with multiple models of the same target system might rather be the norm than the exception. Our point here is that there are different reasons for why inconsistent models are used in science, some of which are taken into account by the different solutions to the problem of inconsistent models in the literature.

In the realist camp, authors like Chakravartty (2010) and Massimi (2018) use the resources of *perspectival realism* (Giere, 2006) to address the problem of multiple models. The idea is, roughly, that the different members of an inconsistent family of models look at their shared target system from different perspectives, under which different features of the target system may emerge. And to properly model the target system from one perspective may require us to make certain assumptions which are inconsistent with those required to model the system from another perspective. But these inconsistencies are innocuous for the *perspectival* realist. Giere (2006), for example, argues that there are ultimately only perspectival facts. Chakravartty (2010) argues instead that if the features that a system exhibits under a given perspective are *dispositional*, we may even learn non-perspectival facts about the target system via perspectival modeling.

In a similar spirit, proponents of *model pluralism* in the social sciences, like Aydinonat (2018) and Grüne-Yanoff and Marchionni (2018), argue that we can use inconsistent families of models to study the same target system, as long as different models in the family are useful for different purposes.

The idea is, roughly, that phenomena in the social sciences can be so complex that they cannot be captured by just one single, overarching model. We need several, possibly mutually inconsistent models, each of which applies to a different context in which the phenomenon occurs. It is the task of the modeler, then, to select the right model for a given context.[11] So, on this approach, a family of mutually inconsistent models functions like a 'bag of tricks' for studying a complex, many-facetted phenomenon.

Each of these accounts provides a way in which a family of mutually inconsistent models can provide coherent information about a shared target system. But note that what these approaches have in common is that they avoid the epistemic problems arising from inconsistency by using different models for different contexts – different perspectives, in the case of perspectival realism, and different purposes in the case of model pluralism. On these accounts, inconsistent families of models essentially provide a *mosaic* account of their target system, where each individual model works like a tile contributing to the bigger, overall mosaic picture. Such a mosaic approach, however, has difficulties when the mutually inconsistent models in question are used to explain one and the same aspect of their target system.

Consider the case of ensembles of climate models discussed by Parker (2010a, 2010b), for example. These models not only share the same target system (viz. a given climate system), but they also aim to account for the same aspect of that system (viz. the climate or weather over a given period). The ensemble methods in fact provide examples of factually inconsistent models (viz. initial condition ensembles) and conceptually inconsistent models (viz. multiple model ensembles) where the mosaic approach fails. Here, mosaic approaches seem to reach their limits: there are no different perspectives or contexts in play, we use the inconsistent models to make predictions about the target system in the same respect in a fixed context. It's precisely these kinds of situations, where multiple mutually inconsistent models are used to account for the same aspect of a shared target system, that we focus on in this paper.

Parker (2010a, 2010b) argues that in the case of ensemble prediction methods, the point is to reduce *uncertainty* concerning the adequate model structure, precise initial conditions, or specific parameter settings. Based on this observation, Parker discusses how and under which conditions, we can obtain probabilistic predictions from ensemble results (Parker, 2010b). But note that Parker's account only provides us with a way of obtaining information about the (probable) *truth* of a claim about the target system. Typically, however, we do not only care about *whether* a result holds, but also *why* it holds – in other words, we want an *explanation* of the results – and this is what the ensemble method is not equipped to (or even intended to) provide.

Take the case of global warming under increased atmospheric CO_2, for example. This result is robust across different climate models, and the ensemble method allows us to infer probabilistic predictions from that (Parker, 2010a, pp. 270–271). But, what the ensemble method does not provide us with is an explanation of the underlying phenomenon.[12] In order to properly understand the prediction, it would be desirable to have a scientific explanation of why we get global warming under increased atmospheric CO_2. The relevant questions are: *Which aspect of the world's climate system necessitates this result? Which laws of nature underlie it?* And so on. Especially when it comes to conceptually inconsistent climate models, at least *prima facie*, the ensemble prediction methods seems ill-equipped to handle these sorts of questions: there are no clear standards for model comparison that can be used with conceptually inconsistent models, i.e. models based on different core causal structures. To be clear, the ensemble method does very well what it's intended to do – to provide predictions about the degree of climate change – our point here is that we cannot use it straightforwardly for our purpose. What we wish to provide is an account of how even what we call conceptually inconsistent models, i.e. models of the same target system that are based on incompatible substantial assumptions, can provide shared scientific explanations of one and the same aspect of their shared target system.

The idea that we shall explore in the following is that even if two models use incompatible substantial assumptions, i.e. they ascribe a different core causal structure to their shared target system,

they can still share a common structure – and this common structure can ground shared information provided by these models. The case study of the following section will provide an example of this. But before we will conduct the study, we will briefly discuss criteria for shared structure among fundamentally different models and how such a shared structure can ground shared information.

How can we show that two models whose fundamental assumptions are incompatible share a common structure? Clearly, concepts like structure preserving maps (isomorphism, partial isomorphism, morphism, etc.) from the debate of the semantic view of theories (see, e.g. Da Costa & French, 1990; French & da Costa, 2000; Landry, 2007) will not be of much help here. These criteria are not only too strict, moreover they require that the models in question are set-theoretic structures: structure preserving maps are functions which require well-defined domains, ranges, and structural components to be preserved. But given the heterogeneity of models assumed in this paper (see p. 5), this is a highly dubious assumption. It is unclear, for example, what a morphism between an equation based model, which is essentially just a system of equations, and a computer simulation model could even look like – there are no underlying sets that can be used as the domains of the functions, the models don't have comparable algebraic structures, and so on. The situation is only made worse by the fact that we're interested in models with fundamentally different substantial assumptions, i.e. models whose mathematical structure can be *very* different. In short, given the heterogeneity of models, it's unreasonable to expect a single, clear criterion for structural similarity across two models – especially if these models are fundamentally different.

Luckily, however, the situation we find ourselves in is not that different from the usual situation in robustness analysis, where we're trying to compare models with the same core causal structure but different auxiliary assumptions (tractability assumptions, idealizations, etc.), cf. Lehtinen (2016, 2018), Weisberg (2006); Weisberg and Reisman (2008) and especially Kuorikoski et al. (2010, pp. 545–549). According to Weisberg (2006, p. 738), the question whether two models exhibit the same core structure in this sense is ultimately a *scientific question* – and as such it can be answered in different ways: We can show, for example, that a model instantiates a structure mathematically, by deriving it as in the case of the differential equation models; or, we can show it empirically, by observing the pattern as in the individual based models. It ultimately depends on the specific assumptions of the models where dealing with and the kind of structure we're interested in.

The main difference to the usual situation in robustness analysis is that in the case of models with fundamentally different assumptions, we're not interested in the core causal structures of the models – those are different by assumption – rather, we're interested in what we might call a shared *peripheral* structure. In the following section, we'll provide an example of a family of trade-models with fundamentally different assumptions about the core mechanisms explaining international trade, which nevertheless share a common peripheral structure about certain trade-patterns, which is captured in a special set of equations (the gravity equations mentioned in the introduction). This peripheral 'gravity structure' is instantiated in different ways in the different models, witnessed by different derivations, based on different assumptions, etc. – but it is still, as we'll argue, a *shared* structure.

Once we've established that there is a shared structure across the inconsistent models, we'll argue that this shared structure grounds shared information provided by the models. We shall argue that the shared structure can ground this information by means of robust theorems of the form:

- '*Ceteris paribus*, if [common (peripheral) structure] obtains, then [robust property] obtains.'

If such a theorem can be established, we argue, every model that shares in the common structure has access to a *structurally similar* explanation of the robust property. In our case study, the shared structure will be a shared gravity structure, which, as we'll show, has been used to account for different scientific claims in the literature – providing us with an example of shared information across mutually inconsistent models.

It might be worth reiterating that what we're interested in here is the problem of inconsistent models: how can mutually inconsistent models, especially models with incompatible substantial assumptions, jointly provide coherent information about a shared target system. The main motivation for us is the fact that scientists apparently *do* use such families of models – our case study will provide an example of this. But at the same time, the case study might be of interest for the debate on robustness analysis, too. As we mentioned above, the literature on robustness analysis typically focuses on cases where the shared structure across models is a shared core causal structure, cf. Kuorikoski et al. (2010). The examples we provide in our case study, instead, appear to be genuine cases of robustness where the relevant structure is not the main causal mechanism of the models. It might thus be interesting to evaluate the case from the perspective of confirmation (using Lehtinen, 2016, 2018), explanatory values (using Ylikoski & Kuorikoski, 2010), or the general role of robustness in economics (using Kuorikoski et al., 2010). But, alas, this work goes beyond the scope of the present paper.

We'd to conclude this section with a remark: in the philosophical literature on inconsistent models, it has been questioned whether the problem of inconsistent models is amenable to philosophical analysis or if it is 'essentially a scientific problem of consistency and coherence in theoretical knowledge' (Morrison, 2011, p. 351). For instance, Morrison voices skepticism concerning the prospects of a philosophical solution to the problem of inconsistent nuclear models in physics. She points out that their success depends precisely on those assumptions that give rise to their inconsistency with models. In light of this, we'd like to point out that we don't aim to provide a general solution of the problem of inconsistent models, which can be applied to any family of mutually inconsistent models and gives us coherent information on what they provide. Rather, we aim to show that in some cases, families of inconsistent models can provide a coherent explanation of a phenomenon via shared structure. The case of the gravity equations is one such example, but we argue that the strategy developed in this paper promises more fruitful applications.

4. Case study: international trade and gravity

In this section, we conduct our case study on the gravity equations in international trade theory. In Section 4.1, we begin with a brief overview of 'classical' trade theories. In Section 4.2, we briefly provide some background on gravity and it's role in international trade theory. In Section 4.3, we discuss structural gravity as an example of shared structure across inconsistent models doing explanatory work. In Section 4.4, we conclude with a philosophical outlook and discussion.

4.1. International trade theory and the gravity equations

Trade theory is a field of economics that encompasses several, often inconsistent models to explain various aspects of trade. Generally speaking, international trade theory aims at answering questions such as:

- Why do countries trade?
- What goods do they trade?
- How much do countries trade and with whom?

The field is intertwined with policy and political implications as different models of trade predict different consequences on countries' economic growth, on income inequality, or on which countries or sectors will benefit from trade and which will suffer. In what follows, we will compare two classical families of models of international trade: models based on comparative advantage and models of specialization. We will provide the main intuitions behind these two groups, with a particular emphasis on the differences that make such models inconsistent with each other.

Before going into the detail, the main ideas behind the two families of model are:

- Models that are based on comparative advantage explain that countries trade with each other because it is favorable for them to export those goods they have a comparative advantage to produce. This means, those goods that they can produce more efficiently than other goods, as compared to another country.
- Models of specialization explain that countries trade with others to take advantage of the inherent benefits of specialization, which allows large-scale production.

Models of comparative advantage.

The concept of comparative advantage lies at the heart of a number of models of international trade, among them the Ricardian model of international trade and the Hecksher-Ohlin model. Even though both models share the concept of comparative advantage, they decline it in different ways. Let us see how.

Ricardo's model

Ricardo's model explains trade on the basis of countries' differences in their technology, which determines different productivity functions across countries.

As an example, suppose that one country is relatively more efficient at producing cloth than wine, as compared to another country; and that the opposite holds for the other country, which is relatively more efficient at producing wine than cloth. The model shows the conditions under which it is convenient for the two countries to trade that good they are more efficient at producing, even only in relative terms.

To see this, let us consider the Ricardian model in some more detail.[13] Suppose that we have only two countries – Home and Foreign – two goods – 1 and 2 – and the only factor of production in the two countries is labor. We also assume that labor is mobile across sectors within countries but not across countries.

First, we denote the productivity of labor a_i at home and a_i^* in the foreign country. The productivity of labor indicates the labor needed per unit of production, for instance, the hours it takes to produce a unit of good 1. The total labor supply at home is L and abroad L^*. Clearly, labor is a limited resource, and this sets a constraint on the total amount of goods that can be produced, i.e. a trade-off between the two goods, which is also called *opportunity cost*. The opportunity cost of good 1 in terms of good 2 represents the units of good 2 that could have been produced with the resources used to produce good 1. A *production possibility frontier* indicates how much of one good can be produced given the production of the other good and vice versa.

In order to establish the actual production of goods at Home, we need to consider the prices of the two goods p_1 and p_2. Assuming that there are no profits over the produced goods, the wage of a worker (w) corresponds to the price per unit of a produced good over the unit of labor, i.e. $w_1 = p_1/a_1$ and $w_2 = p_2/a_2$. Wages in sector 1 are higher than in sector 2 when $p_1/p_2 > a_2/a_1$. Given that workers prefer to maximize their wages, production will specialize in sector 1 if $p_1/p_2 > a_2/a_1$; it will specialize in sector 2 if $p_1/p_2 < a_2/a_1$; and an economy will produce both goods if $p_1/p_2 = a_2/a_1$.

Suppose that Home has a comparative advantage in producing good 1, i.e. $a_1/a_2 < a_1^*/a_2^*$. This implies that at Home the relative price of good 1 is lower than that abroad, at least when the two countries do not trade.

Let us now consider what happens when countries engage in international trade. Given that Home relative price of good 1 is lower than abroad, then it is efficient for Home to trade Good 1 on the world market and for Foreign to buy it and vice versa for Good 2. Thus, we can focus on the situation where $p^a < p < p^{a*}$, where p is the world price. In this case, Home will fully specialize in good 1 and trade it at the relative price $p > p^a$; a similar line of reasoning applies to Foreign for good 2. This allows us to conclude that both countries are better off under free trade. In other words, by producing good 1 and trading it for good 2, Home can have more units of good 2 than it would have under autarky. The same applies to Foreign for good 1.

We have already seen that Ricardo's model makes a number of assumptions, such as that there is only one factor of production and labor mobility across sectors. Beyond these, it assumes perfect competition and constant returns to scale. It also assumes that tastes are identical and homogeneous. These assumptions determine a number of consequences, such as that each country will only produce that good in which it has a comparative advantage. It also implies that everyone within one country will benefit from international trade. Many of these consequences are clearly not observed in the real world: countries do not specialize in just one good, and not everyone benefits from international trade. Nevertheless, Ricardo's model is considered to be important as the basic idea of the model, i.e. that a country will trade the goods in which its productivity is relatively high, is still confirmed by the evidence. Krugman, Melitz, and Obstfeld, for instance, summarize the empirical evidence on the Ricardian model this way:

> 'Does the Ricardian model make accurate predictions about actual international trade flows? The answer is a heavily qualified yes. Clearly there are a number of ways in which the Ricardian model makes misleading predictions. […] In spite of these failings, however, the basic prediction of the Ricardian model – that countries should tend to export those goods in which their productivity is relatively high – has been strongly confirmed by a number of studies over the years.' (Krugman et al., 2012, p. 45)

The Hecksher-Ohlin model

The H-O model also builds on comparative advantage, even though it is a different kind of comparative advantage than in the Ricardian model. In the H-O model, unlike in Ricardo's model, there are no differences in the production function between countries, i.e. in their technologies. The model assumes that trading depends on other factors of productivity beyond labor, i.e. land and capital. The main difference between countries is in their relative endowment of factors of production, such as land, capital, and labor.

Let us look at the H-O model in some more detail.[14] Suppose that there are two countries – Home and Foreign – two goods – 1 and 2 – and two factors of production – labor (L) and capital (K). In the model, Home and Foreign differ in the relative endowment of factors of production. Home is *relatively labor abundant* if it has a ratio of labor to capital that is higher relative to that of the other country. In this case, Foreign is relatively capital abundant:

$$L/K > L^*/K^* \tag{1}$$

Goods differ in their *relative factor intensity*, which is the intensity with which factors are employed in the production of goods, i.e. the ratio of resources needed for production. For instance, good 1 is relatively labor intensive if it requires more labor to capital than good 2 and, conversely, good 2 is relatively capital intensive:

$$L_1/K_1 > L_2/K_2 \tag{2}$$

Other assumptions of the model are that the factors of production are mobile across sectors within countries, but not across countries, and that tastes are identical and homogeneous.

Under these assumptions, the H-O model shows that each country will export the good that uses its relatively abundant factor intensively. To see this, let us focus on our example where Home is relatively labor abundant, while Foreign is capital abundant. Recall also that good 1 is relatively labor intensive and good 2 relatively capital intensive. This implies that Home produces more of good 1 than good 2, as compared to Foreign, and that the price of good 1 at home is lower than abroad and vice versa.

When the two countries start trading, the relative price of the goods tend to converge. This way, for the same relative price, Home will have a larger relative supply of good 1 than Foreign.

Since the model assumes that tastes are identical and homogeneous, the demand curve will be the same for both countries. With trade, a new relative price of good 1 will settle somewhere between the pre-trade prices. Since the relative price of good 1 increases for Home as compared

to pre-trade levels, Home will export good 1 and Foreign will export good 2. This way, Home special-izes more in good 1 and exports it in exchange for more units of good 2 than in the absence of trade. The opposite holds for Foreign, which specializes more in good 2 and exports it in exchange for more units of good 1 than in the absence of trade.

This way, the H-O model shows that each country will export that good that uses intensively the factor which is relative abundant in the country. In other words, the model explains trade on the basis of the proportion of factor availability and the relative intensity with which factors are used in producing goods.

Attempts to use the H-O model to explain empirical data have been keeping international trade theorists busy, starting at least from the work of Leontief (1953) until the present time. Authors typi-cally agree that the assumption that technologies are the same across countries is problematic, and that international trade is largely driven by differences in the productivity of labor. For instance, according to Feenstra:

> 'The H-O model is hopelessly inadequate as an explanation for historical or modern trade pattern unless we allow for technological differences across countries. For this reason, the Ricardian model is as relevant today as it has always been.' (Feenstra, 2015, p. 1).

Other authors praise the empirical merits of the H-O model especially for its ability to explain pattern of international trade in specific contexts, such as trade between developing and devel-oped countries (on this, see, e.g. Krugman et al., 2012; Leamer, 2012).[15] Leamer's (2012) view on the H-O model is more nuanced. According to him, attempts to 'try to grasp the data tightly to determine if trade, technologies, and factor endowments conform with the HOV [Heckscher-Ohlin-Vaneck] predictions have found what seem like serious inaccuracies in the HOV model'. On the other hand, however, studies that 'take the central prediction of the HO framework to be that the structure of trade is correlated (in some way) with relative factor supplies (measured some how) […] are much more favorable to the HO framework' (Leamer, 2012, pp. 121–122).

Before we move on to models of specialization, it is worth focusing more closely on Leamer's view of the role of economic models, as it is different than the role emphasized. In the Ohlin Lectures that he delivered in 2009, which served as the basis for his 2012 book 'The Craft of Economics', Leamer argues that economic models should be appraised in terms of their fruitfulness. In his own words: 'Our goal as economists should be usefulness and insights, not necessarily validity. A model can be valid, but entirely useless or worse misleading. And […] a model/framework/thought can be insightful, even if it isn't valid.' (Leamer, 2012, p. 16). He also says: 'I will argue that the primary goal of economic modeling is *not to predict or to explain*, but instead to assist in the design of gov-ernment market interventions.' (p. 6, italics added).

Leamer distinguishes between theoretical and empirical models and stresses that *the great econ-omist* is the one who is able to find a balance between the two. An important part of the economists' work, which according to Leamer is often neglected, is to move from a theoretical model to an empirical model by means of specifying the former so that it can serve as a basis for empirical testing. In this respect, he claims that the assumptions used in theoretical models are 'abstractions to make a point, not abstractions purposefully chosen to approximate the real world.' (Leamer, 2012, p. 108)

Leamer's view on the relation between theoretical models and empirical models would require an extensive detour, but the main point that is relevant to us here is that he claims that the purpose of theoretical economic models is not to explain phenomena.[16] We leave it open if Leamer is right on *the* purpose of economic models and we bracket the substantial question of what exactly is that makes theoretical models helpful for policy interventions. We are interested in a different, narrower question, which is: can inconsistent models explain a given phenomenon and, if so, under what con-ditions? We argue, as it will become clearer in the next sections, that one way in which inconsistent models can explain is by means of their shared structure.

Models of specialization

Models that explain trade on the basis of specialization give up several assumptions that are present in the two previous models and in particular they abandon constant returns to scale in favor of increasing returns to scale. Unlike models of comparative advantage, which show that *differences* between countries explain why they import or export a certain product, models of specialization explain why even countries that are similar to each other, in terms of factors of production, can still gain from trade.

Intuitively, it is easy to see the logic underlying models of specialization: suppose that two countries have the same factors of production and that each of them produces two goods for the domestic market. Assume also that these goods are in industries where we have economies of scale. Then, the supply curve for both goods is forward-falling – the more units are produced, the lower the price, as the average cost of production falls. In the absence of trade, each country has only access to their domestic market. Given that a domestic market might be too small to reap the benefits of a scale economy, countries that open up to trade can gain from it.[17]

It is also important to emphasize that in models of specialization, the potential benefits from trade do not derive from comparative advantage, since countries are endowed with the same factors of production by assumption. Models of specialization typically also relax the assumption of perfect competition in favor of monopolistic competition. Under perfect competition, countries produce homogeneous goods, which means, there are no differences in the quality of goods they produce. By contrast, under monopolistic competition, countries differentiate the production of similar products. This way, models of specialization gained support and credibility as they are able to explain the high and increasing proportion of trade of similar products between similar countries worldwide (Krugman, 2009, p. 561).[18]

Notice that, by contrast, comparative advantage models do not explain intra-industry trade, as they predict that countries with similar resources will not trade with each other. But given that the actual world shows high level of intra-industry trade, to many economists this seemed like a knockdown argument against comparative advantage models in favor of models of specialization:

> 'These theories [Ricardo and H-O] provided good explanations of the trade patterns in the first half of the 20th century. But as many researchers began to observe, comparative advantage seemed less relevant in the modern world. Today, most trade takes place between countries with similar technologies and similar factor proportions; quite similar goods are often both exported and imported by the same country. […] Under such circumstances, how could intra-industry trade be explained? The traditional view, that a given country would have a comparative advantage in terms of technology or factor endowments when producing a particular type of textile, seems far-fetched as an explanation.' (The Royal Swedish Academy of Science, 2008, p. 1).

This view, however, is not unanimously shared by economists. According to Leamer (2012), for instance, it would be a mistake to abandon the H-O model, as the model still helps us illuminating patterns of international trade in ways that models of specialization do not (see, on this, also Evenett & Keller, 2002). To see Leamer's argument, let us focus on a specific point, which concerns the role of distance in trade. Models of specialization include distance as a determinant of trade that, simplifying things, affects the price of goods this way: the larger the distance that goods have to travel, the higher their price. In the H-O models, by contrast, there are no costs of commerce – the only role of distance is that labor and capital cannot move across countries (Leamer, 2012).[19]

Nevertheless, according to Leamer, the role of distance has been overestimated in the literature. In Blum and Leamer (2004), Leamer shows that agricultural goods are the least affected by distance whereas manufacturing goods are negatively affected by distance. The main upshot is that the H-O model cannot be neglected because, even when accounting for the role of distance, differences in resource endowments still matter to international trade. Another reason why Leamer is not willing to discard the H-O model is that he claims it can shed light on labor market effects of trade via the Stolper-Samuelson model, and, relatedly, affects trade on income inequality (Blum & Leamer, 2004). In Leamer's words:

'In conclusion, […] the HO model continues to provide surprising insights. Not only is it useful as a theory. It accurately explains many prominent features of the patterns of international trade, and it is an essential ingredient in any study of the impact of globalization on the U.S. work force.' (Leamer, 1995).

Inconsistent models

Let us recapitulate where we stand. The previous two families of models, i.e. comparative advantage models and models of specialization, explain trade in diametrically opposite ways. The idea of comparative advantage is that countries trade with each other because of their *differences*. It is either because they are endowed with different resources, or because of their different technology, that it is efficient for them to exchange those goods they can produce comparatively better than others. Models of specialization, by contrast, do not need any differences between countries in terms of factors of production or technologies. Countries that are similar to each other with respect to their endowment of resources gain nevertheless from trade because of the benefits of expanding markets.

The two groups of models make considerably different assumptions to get to their results: perfect competition and constant returns to scale on the one hand and imperfect competition and increasing returns to scale on the other hand. Even though these two families are based on different assumptions and provide 'rival' explanations, economists usually do not claim that there is *one* model of international trade that is in absolute terms better than the others (Krugman, 2011). These models are usually presented as models each of which highlight some important features of the phenomena.

Krugman (2011), for instance, compares the two previous approaches side by side. According to him, the history of international trade can be divided into eras where comparative advantage 'ruled' as against others where specialization dominated trade flows. At the same time, he also claims that it would be a mistake to think that only comparative advantage explains the whole story of international trade even in those phases dominated by comparative advantage.

In Krugman's words:

'[E]ven during comparative advantage eras increasing returns in the form of localized external economies plays a significant role. […] [I]ncreasing returns […] have historically played an important role in world trade even in eras when comparative advantage seems to reign supreme.' (Krugman, 2011, p. 51).

4.2. Traditional gravity

In its simplest form, the gravity model states that the trade flow between two countries is proportional to (the product of) their 'economic masses' and inversely proportional to their 'distance.' This idea gives rise to the simple gravity equation:

$$X_{i,j} = g\frac{X_i X_j}{d_{i,j}}, \tag{3}$$

where $X_{i,j}$ is the bilateral trade flow between i and j, g is a 'gravitational constant' of proportionality, X_i and X_j are the 'economic masses' of i and j, and $d_{i,j}$ is the 'distance' between the two countries. These variables can be instantiated in various ways, but in the most straight-forward simple gravity models X_i, X_j are simply the GDP's of i and j and $d_{i,j}$ their physical distance.

Simple gravity entered economics as an econometric data model in the studies of Tinbergen (1962). The model has been quite successful empirically and found several policy applications, cf. Van Bergeijk and Brakman (2010), Head and Mayer (2014), Anderson (2011), and Kabir et al. (2017). A striking example of the empirical success of the gravity equations is the study of Havrylyshyn et al. (1991), who used gravity estimates to predict with surprising accuracy how trade patterns would be changed by the fall of the Iron Curtain, see Van Bergeijk and Brakman (2010, p. 7).

Despite their empirical success and usefulness in policy applications, economists were initially skeptical of gravity because of its perceived lack a solid theoretical (microeconomic) foundation.

Leamer and Levinsohn, e.g. lament: 'The gravity models are strictly descriptive. They lack a theoretical underpinning so that once the facts are out, it is not clear what to make of them' (Leamer & Levinsohn, 1995, p. 1387).

This skepticism, however, was ultimately dispelled by a plenitude of derivations of gravity equations in a wide range of different trade models. To name just a few:

- Eaton and Kortum (2002) derive gravity in a Ricardian model of trade;
- Deardorff (1998) gives a derivation of gravity equations that are consistent with the assumptions of Heckscher-Ohlin models;
- Anderson (1979) and Anderson and Van Wincoop (2003) provide a derivation in models of comparative advantage.

A particular driving force for theoreticians to derive gravity was to avoid their favorite model of international trade being falsified by the empirically robust data encoded in the gravity equations (Deardorff, 1998, p. 7). This worry, however, is dispelled by derivations in models of essentially all major trade theories. As Deardorff concludes: 'because the gravity equation appears to characterize a large class of models, its use for empirical tests of any of them is suspect' (Deardorff, 1998, p. 21).

For our purposes, however, the fact that all these mutually inconsistent models derive gravity equations holds philosophical promise: it supports the hypothesis that there is a shared structure among these models that can be used to provide a common explanation of (empirically observed) trade-flow patterns. We shall investigate this hypothesis in the remainder of this section, for this purpose, we'll take a closer look at the so-called *structural* gravity model.

4.3. Structural gravity

The structural gravity model was developed by Anderson and Van Wincoop (2003, 2004) as a solution to McCallum's infamous 'border puzzle' for gravity (McCallum, 1995). This puzzle essentially consists in the fact that simple gravity models over-predict trade between U.S. states and Canadian provinces taken as units of trade-flow analysis. A simplified summary of the problem is this: since U.S. states and adjacent Canadian provinces are physically close, gravity predicts high trade volumes between them; in reality, however, we observe an apparent trade-destroying fact at the U.S.-Canadian border; the states and provinces disproportionally trade among themselves rather than across the border. The only way for gravity models to accommodate this seemed to be by means of *ad hoc* dummy variables, which makes for a theoretically unsatisfactory solution.[20]

To address the puzzle, Anderson and van Wincoop focus on what they've coined 'multilateral resistance parameters' in gravity. In a model of specialization with identical homothetic preferences, approximated by a utility function with constant elasticity of substitution (CES) σ, Anderson and van Wincoop derive the gravity equation (4) subject to conditions (5) and (6), which implicitly define the resistance terms Π_i and P_j respectively:

$$X_{i,j} = \frac{X_i X_j}{X_w} \left(\frac{d_{i,j}}{P_j \Pi_j} \right)^{1-\sigma} \tag{4}$$

$$(\Pi_i)^{1-\sigma} = \sum_j \left(\frac{d_{i,j}}{P_j} \right)^{1-\sigma} \frac{X_j}{X_w} \tag{5}$$

$$(P_j)^{1-\sigma} = \sum_j \left(\frac{d_{i,j}}{P_j} \right)^{1-\sigma} \frac{X_i}{X_w} \tag{6}$$

In these equations, X_w is the global 'economic mass' and Π_j is j's outward multilateral resistance, while P_j is j's inward multilateral resistance, both measuring i's and j's ease of market access, intuitively.

Something that's particularly interesting to note for our purposes is that by implicitly defining the resistance terms Π_i and P_j via Equations (5) and (6), Anderson and van Wincoop tie their gravity equation to the internal structure of their model – whence the name 'structural gravity.' In fact, as Anderson (2011, p. 142) points out, we can solve (6) and (7) – up to a multiplicative constant – from X_i, X_j and $d_{i,j}^{1-\sigma}$ for all i,j.

Anderson and van Wincoop go on to empirically estimate the parameters of their model for the case of U.S. states and Canadian provinces, which allows them to solve McCallum's border puzzle. They find that, via their resistance terms, the international border has a trade-reducing effect of around 20–50% (Anderson & Van Wincoop, 2003).

The derivation of structural gravity by Anderson and van Wincoop is what's called a 'demand side' derivation, since its based on assumptions about demand (homothetic preferences with CES). As it turns out, however, structurally identical, *isomorphic* gravity equations can also be derived from the supply side. Such a derivation can be provided in Eaton and Kortum's Ricardian trade model (Eaton & Kortum, 2002). In their model, it is determined probabilistically whether a country is likely to provide a certain good to another country based on a so-called *Fréchet* distribution with a variation parameter θ that is common to all countries. The equations that can be derived in Eaton and Kortum are (see Yotov et al., 2016, pp. 57–59 for a detailed derivation):

$$X_{i,j} = \frac{X_i X_j}{X_w} \left(\frac{d_{i,j}}{P_j \Pi_j} \right)^{\theta} \tag{7}$$

$$(\Pi_i)^{\theta} = \sum_j \left(\frac{d_{i,j}}{P_j} \right)^{\theta} \frac{X_j}{X_w} \tag{8}$$

$$(P_j)^{\theta} = \sum_j \left(\frac{d_{i,j}}{P_j} \right)^{\theta} \frac{X_i}{X_w} \tag{9}$$

Note that the only difference between Equations (4)–(6) and (7)–(9) is that the variation parameter θ is substituted for $1 - \sigma$, which allows us to argue that the two sets of equations express one and the same structure. This structure can be represented by the following parametized equations:

$$X_{i,j} = \frac{X_i X_j}{X_w} \left(\frac{d_{i,j}}{P_j \Pi_j} \right)^{\mathcal{X}} \tag{10}$$

$$(\Pi_i)^{\mathcal{X}} = \sum_j \left(\frac{d_{i,j}}{P_j} \right)^{\mathcal{X}} \frac{X_j}{X_w} \tag{11}$$

$$(P_j)^{\mathcal{X}} = \sum_j \left(\frac{d_{i,j}}{P_j} \right)^{\mathcal{X}} \frac{X_i}{X_w} \tag{12}$$

If we set the parameter $\mathcal{X}$ to $1 - \sigma$, we obtain Equations (4)–(6); if, instead, we set $\mathcal{X}$ to θ, we obtain Equations (7)–(9).

What conclusions can we draw from this? First, we wish to argue that the two mutually inconsistent models (cf. Table 1), Anderson and van Wincoop's model of specialization on the one side and Eaton and Kortum's Ricardian model on the other side, share a common *gravity structure*, expressed by the parametized equations (10)–(12). This structure is instantiated in different ways in the two models – on the demand side in Anderson and van Wincoop's model and on the supply side in Eaton and Kortum's model – witnessed by different derivations, but the mathematical structure is the same.

This means that the two models have in principle access to structurally similar explanations of empirical phenomena if they are based on the structural gravity equations. One could, for

example, try to provide a Ricardian answer to McCallum's border puzzle, which is structurally identical to Anderson and van Wincoop's solution, by empirically estimating the model parameters in (7)–(9) – though this has not been done to the best of our knowledge.

We do, however, find economic studies in the literature that make use of the shared structure between these mutually inconsistent models to provide theoretical explanations. We mention two:

- Arkolakis et al. (2012) introduce a welfare formula that, under certain assumptions, arises in a large class of trade models, including the models of Anderson and van Wincoop and Eaton and Kortum. A central result of Arkolakis et al. is that, in the setting of the paper, the gains from trade are invariant under the parameters of the isomorphic gravity equations (Arkolakis et al., 2012, pp. 117–118) (cf. also Yotov et al., 2016, p. 3 on this point). Very roughly, we can reformulate this result as a robust theorem in Weisberg's sense as follows:

  Ceteris paribus, if [same gravity parameters], then [same gains from trade].

  Here the 'ceteris paribus' clause covers the model structure needed for the welfare formula to hold. Arkolakis et al. use this robust result to argue that new trade models don't provide new insights into the question of gains from trade: since both old and new models are in the class of models for which the theorem holds, the welfare gains in these models are the same. That is, by means of their robust theorem, Arkolakis et al. extract shared information from a class of mutually inconsistent models of the same phenomenon.
- The study of Allen et al. (2020) provides another example. The authors of this study show that in a large class of models (again including the models of both Anderson and van Wincoop as well as Eaton and Kortum), the existence of equilibria and the counterfactual predictions concerning trade flows, incomes, and prices purely depends on the parameters of the isomorphic gravity equations:

  Ceteris paribus, *if* [*same gravity parameters*], *then* [*same counterfactual predictions*].

  Again, it seems that what Arkolakis et al. are doing here is to extract information from a large class of mutually inconsistent models by means of a shared gravity structure.

Table 1. The assumptions underlying the three models presented in the section.

		Inconsistent assumptions between comparative advantage models and models of specialization	Inconsistent assumptions between Ricardo's model and H-O model
Comparative advantage	Ricardo	• Constant return to scale	• Countries differ in production technology (productivity)
		• Perfect competition	• Labor is mobile across sectors within countries
		• Two countries, home and foreign • Two goods, 1 and 2 • Tastes are identical homogeneous	
	H-O model	• Constant return to scale	• Countries differ in their factor endowment
		• Perfect competition	• No differences in technology across countries
		• Two countries, home and foreign	• Factors of production are mobile across sectors within countries
		• Two goods, 1 and 2 • Tastes are identical homogeneous	
Models of specialization		• Economies of scale	
		• Monopolistic competition	
		• Each firm produces a good that is similar to but differentiated from the goods that other firms in the industry produce	

We claim that these two studies provide clear examples of economic work where information is extracted from mutually inconsistent models (we've focused on Ricardian and New Trade models) via a shared structure (the structural gravity equations). From the perspective of the inconsistent models problem, our story ends here: we have an instance of shared information across mutually inconsistent models by means of a shared structure. We leave an evaluation of the case study from the perspective of robustness analysis, especially in the context of Kuorikoski et al.'s approach to economic modeling as robustness analysis, for future work.

5. Conclusion

In this paper, we have focused on conceptually inconsistent models, i.e. models with different substantial assumptions, of the same target system. These models differ from each other along multiple lines, ranging from their variables to their relations. Inconsistent models represent the same target system in mutually incompatible ways. As we have seen, this inevitably raises a number of questions: How can inconsistent models jointly provide a coherent explanation of the same phenomenon? And in case they lead to the same result, what can we conclude from that?

We argued that the 'quick' reaction of discarding the problem as a case of mere scientific disagreement is too fast. As we have discussed in this paper, it is often the case that scientists work with families of inconsistent models, and not with the purpose of finding *the* right model, or of dismissing the wrong models in favor of better models. Rather, they often use inconsistent models simultaneously as a source of information about the system of interest. We have provided examples from physics, climate science and economics.

But even though scientists are working with inconsistent models, the question remains: what is the rationale behind this practice? In this paper, we have observed that some of the standard answers provided in the literature fall short, at least in some instances where inconsistent models are used. In particular, the ideas from perspectival realism and model pluralism do not apply to cases where inconsistent models refer to the *same* phenomenon in the *same* target system. We have argued that one possible way in which families of inconsistent models jointly explain a given phenomenon in such situations is by sharing an underlying structure, which is responsible for the phenomenon of interest. If we can show that the different representations the models display share an underlying structure that determines the result, then we have a reason to claim that the models jointly explain the same phenomenon *via* this structure. This way, we can go beyond the differences that the models exhibit at the surface level.

International trade theory is a case in point: models of international trade make different assumptions and use different variables to describe flows of international trade. And yet, economic practice suggests that if we want to understand a complex phenomenon such as international trade, we cannot rely on just one single overarching model. We need several, different, possibly inconsistent models to tackle the problem at issue. Moreover, inconsistent models of international trade are all compatible with the data encoded in the gravity equations. Therefore, they cannot be discarded on the basis of the (same) observational evidence. By showing that they converge on the same result in virtue of their underlying structure, we can understand how they can be used together in a meaningful way. In this paper, we proposed the use of structural gravity in the studies of Arkolakis et al. (2012) and Allen et al. (2020) as examples of precisely this.

In doing so, we have shown a possible way in which inconsistent models can be used together and jointly explain a result. If we would only focus on the inconsistencies across them, and consider them as cases of scientific disagreement, we would loose an important part of what scientists can do with them.

We'd like to conclude with a brief remark on a promising connection between the approach we've endorsed in this paper and *structural realism* (Ladyman, 2019; Ladyman & Ross, n.d.). A natural view for a structural realist to take is that what makes a model a good model of its target system is that it accurately captures the *structure* of reality. Adopting such a view would give us an interesting line of

support for the answer to the problem of inconsistent models that we've described in this paper. The idea would be that a family of mutually inconsistent models can provide an explanation of a phenomenon in the real world by deriving that result from a shared structure that latches on to the actual structure of the target system in reality. In fact, the case study we've conducted in this paper, might provide an interesting case in point here, especially in light of the general suitability of structural realism to provide an account of economics (Ross, 2008). Obviously, much more would need to be said to turn this into a philosophically acceptable account, this is just an open question to develop in future research.

Notes

1. In this paper, we don't focus on models that are inconsistent in the sense of one model containing contradicting assumptions. For an overview of the literature on inconsistent science in this sense, see Bueno and Vickers (2014). In the following, when we speak of 'inconsistent models', we always mean 'mutually inconsistent models'.
2. For an overview of the history of gravity in trade theory, see Head and Mayer (2014, Section 1.2).
3. See Yotov et al. (2016) for an overview of structural gravity and, in particular, its applications in policy.
4. For an overview of the vast philosophical literature on scientific models, see Frigg and Hartmann (2018).
5. For a more detailed discussion of the concept of a target system, see Elliott-Graves (2014).
6. See Basso et al. (2017) for an overview of scientific models' epistemic functions.
7. For a discussions of the implications of this view, see Frigg and Hartmann (2018, Section 4.2) and the essays in Morgan and Morrison (1999).
8. For an overview of the different views on the ontology of models in the literature, see Frigg and Hartmann (2018, Section 2).
9. The question in virtue of what two models based on incompatible assumptions can be considered models of the same target system is intriguingly complex, but we shall not address it in this paper. For more on this question, see the discussion of *the problem of style* in Frigg (2006); Frigg and Nguyen (2017).
10. But see Easwaran and Fitelson (2015) and Hughes (2017) for a critical discussion.
11. In his book 'Economic Rules', Rodrik (2017) provides an illuminating example of model pluralism in economics. Consider a question such as: 'Does a reduction in the government fiscal deficit hamper or stimulate economic activity?' (Rodrik, 2017, p. 18). According to a standard macroeconomic model, the answer to the question is that fiscal cuts hamper economic activity. But, the model that shows this result makes a number of assumptions, such as the particular monetary policy a government implements. Under a different monetary policy, a different answer would follow. Therefore, in order to explain the relation between fiscal cutbacks and economic activity, we need a family of models that make different assumptions among which to select the one that applies to the context of interest.
12. I fact, the ensemble method does not aim to do that (see also below). There are well-known explanations of why global warming occurs via the greenhouse effect, for an overview, see, e.g. Jain (1993). But our point here is different: it's just that the ensemble method as discussed by Parker does not allow us to extract explanatory information from families of mutually inconsistent models.
13. In what follows, we are mainly following Feenstra (2015); the interested reader can refer to Feenstra and Taylor (2008), and Krugman et al. (2012).
14. In what follows, we are mainly drawing on Krugman et al. (2012); the interested reader can refer to Feenstra (2015) and Feenstra and Taylor (2008)
15. On this note, according to Leamer: 'The HO framework successfully explains why Latin American tropical countries with abundant natural resources and abundant workforces with low educational attainment export coffee and bananas and oil and minerals, while Asian countries with less abundant natural resources export apparel and footwear.' (Leamer & Stern, 2006, pp. 5–6)
16. See Ross (2014) for a discussion of Leamer's view on theoretical models in macroeconomics in particular.
17. Notice that models of specialization do not rely on increasing returns throughout the range of the production function. The key relationship is a minimal efficient scale that is larger than the extent of the domestic market.
18. To be more precise, models of specialization predict international trade even under perfect competition at the firm level. See, on this, the distinction between internal economies of scale and external economies of scale, as explained for instance in Krugman et al. (2012).
19. On this aspect:

   'An HO model has a very peculiar external geography with countries infinitely far apart as far as the migration of factors of production is concerned (including capital as well as labor), but infinitesimally close to each other as far as the cost of commerce is concerned.' (Leamer, 2012, p. 108)

20. See Head and Mayer (2014), Anderson (2011), and Yotov et al. (2016) for a more detailed account of the development of structural gravity.

Acknowledgments

Disclosure statement

Notes on contributors

Chiara Lisciandra is Assistant Professor at the University of Groningen, Faculty of Economics and Business. She is an elected Board Member of the International Network for Economic Methodology (INEM), a Board Member of the Centre for Philosophy, Politics and Economics at the University of Groningen, and an Associate Member of the Centre for the Philosophy of the Social Science, University of Helsinki. She has published extensively on norms in a wide range of scientific journals, such as *Philosophical Studies*, *Synthese and Politics*, *Philosophy & Economics*. Her primary research interests are in general philosophy of science, philosophy of economics and social philosophy. URL: https://www.chiaralisciandra.com

Johannes Korbmacher is Assistant Professor at the Department of Philosophy and Religious Studies of Utrecht University. He's a mathematical philosopher who mainly works at the intersection of logic, epistemology, and metaphysics. http://jkorbmacher.org/.

References

Allen, T., Arkolakis, C., & Takahashi, Y. (2020). Universal gravity. *Journal of Political Economy*, *128*(2), 393–433. https://doi.org/10.1086/704385

Anderson, J. E. (1979). A theoretical foundation for the gravity equation. *The American Economic Review*, *69*(1), 106–116.

Anderson, J. E. (2011). The gravity model. *Annual Review of Economics*, *3*(1), 133–160. https://doi.org/10.1146/annurev-economics-111809-125114

Anderson, J. E., & Van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review*, *93*(1), 170–192. https://doi.org/10.1257/000282803321455214

Anderson, J. E., & Van Wincoop, E. (2004). Trade costs. *Journal of Economic Literature*, *42*(3), 691–751. https://doi.org/10.1257/0022051042177649

Arkolakis, C., Costinot, A., & Rodríguez-Clare, A. (2012). New trade models, same old gains? *American Economic Review*, *102*(1), 94–130. https://doi.org/10.1257/aer.102.1.94

Aydinonat, E. (2018). The diversity of models as a means to better explanations in economics. *Journal of Economic Methodology*, *25*(3), 237–251. https://doi.org/10.1080/1350178X.2018.1488478

Basso, A., Lisciandra, C., & Marchionni, C. (2017). Hypothetical models in social science. In *Springer handbook of model-based science* (pp. 413–433). Springer.

Bergstrand, J. H. (1985). The gravity equation in international trade: Some microeconomic foundations and empirical evidence. *The Review of Economics and Statistics*, *67*(3), 474–481. https://doi.org/10.2307/1925976

Blum, B., & Leamer, E. (2004). Can FTAA suspend the law of gravity and give Americas higher growth and better income distributions. In T. Estevadeordal, D. Rodrick, A. Taylor, & A. Velasco (Eds.), *Integrating the Americas: Ftaa and beyond* (pp. 539–572). Harvard University Press.

Bueno, O., & Vickers, P. (2014). Is science inconsistent? *Synthese*, *191*(13), 2887–2889. https://doi.org/10.1007/s11229-014-0463-9

Cartwright, N. (1983). *How the laws of physics lie*. Oxford University Press.

Chakravartty, A. (2010). Perspectivism, inconsistent models, and contrastive explanation. *Studies in History and Philosophy of Science Part A*, *41*(4), 405–412. https://doi.org/10.1016/j.shpsa.2010.10.007

Da Costa, N. C., & French, S. (1990). The model-theoretic approach in the philosophy of science. *Philosophy of Science*, *57*(2), 248–265. https://doi.org/10.1086/289546

Deardorff, A. (1998). Determinants of bilateral trade: Does gravity work in a neoclassical world? In J. A. Frankel (Ed.), *The regionalization of the world economy* (pp. 7–32). University of Chicago Press.

Easwaran, K., & Fitelson, B. (2015). Accuracy, coherence, and evidence. *Oxford Studies in Epistemology*, *5*, 61–96. https://doi.org/10.1093/acprof:oso/9780198722762.001.0001

Eaton, J., & Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, *70*(5), 1741–1779. https://doi.org/10.1111/ecta.2002.70.issue-5

Einstein, A., & Infeld, L. (1938). *The evolution of physics: The growth of ideas from early concepts to relativity and quanta*. Cambridge University Press.

Elliott-Graves, A. (2014). *The role of target systems in scientific practice* [Doctoral dissertation, University of Pennsylvania]. https://repository.upenn.edu/edissertations/1267/

Evenett, S. J., & Keller, W. (2002). On theories explaining the success of the gravity equation. *Journal of Political Economy*, *110*(2), 281–316. https://doi.org/10.1086/338746

Feenstra, R. (2015). *Advanced international trade: Theory and evidence*. Princeton university press.

Feenstra, R., & Taylor, A. (2008). *International economics*. Worth.

French, S. (2010). Keeping quiet on the ontology of models. *Synthese*, *172*(2), 231–249. https://doi.org/10.1007/s11229-009-9504-1

French, S., & da Costa, N. (2000). Theories, models and structures: Thirty years on. *Philosophy of Science, Supplement: Proceedings of PSA*, *67*(1), 116–127. https://doi.org/10.1086/392813

Frigg, R. (2006). Scientific representation and the semantic view of theories. *Theoria*, *21*(1), 49–65.

Frigg, R., & Hartmann, S. (2018). Models in science. In E. Zalta (Ed.), *Stanford encyclopedia of philosophy* (Summer 2018). Metaphysics Research Lab, Stanford University.

Frigg, R., & Nguyen, J. (2017). Models and representation. In *Springer handbook of model-based science* (pp. 49–102). Springer.

Giere, R. N. (1988). *Explaining science: A cognitive approach*. University of Chicago Press.

Giere, R. N. (2006). *Scientific perspectivism*. University of Chicago Press.

Grüne-Yanoff, T., & Marchionni, C. (2018). Modeling model selection in model pluralism. *Journal of Economic Methodology*, *25*(3), 265–275. https://doi.org/10.1080/1350178X.2018.1488572

Havrylyshyn, O., Pritchett, L., & Mundial, B. (1991). *European trade patterns after the transition* (World Bank Pre Working Paper Series 748).

Head, K., & Mayer, T. (2014). Gravity equations: Workhorse, toolkit, and cookbook. In *Handbook of international economics* (pp. 131–195). Elsevier.

Hughes, M. (2017). Epistemic inconsistency and categorical coherence: A study of probabilistic measures of coherence. *Synthese*, *194*(8), 3153–3185. https://doi.org/10.1007/s11229-016-1105-1

Jain, P. (1993). Greenhouse effect and climate change: Scientific basis and overview. *Renewable Energy*, *3*(4–5), 403–420. https://doi.org/10.1016/0960-1481(93)90108-S

Kabir, M., Salim, R., & Al-Mawali, N. (2017). The gravity model and trade flows: Recent developments in econometric modeling and empirical evidence. *Economic Analysis and Policy*, *56*, 60–71. https://doi.org/10.1016/j.eap.2017.08.005

Knuuttila, T. (2005). Models, representation, and mediation. *Philosophy of Science*, *72*(5), 1260–1271. https://doi.org/10.1086/508124

Knuuttila, T. (2011). Modelling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science Part A*, *42*(2), 262–271. https://doi.org/10.1016/j.shpsa.2010.11.034

Krugman, P. (2009). The increasing returns revolution in trade and geography. *American Economic Review*, *99*(3), 561–571. https://doi.org/10.1257/aer.99.3.561

Krugman, P. (2011). Increasing returns in a comparative advantage world. In R. Stern (Ed.), *Comparative advantage, growth, and the gains from trade and globalization: A festschrift in honor of Alan V Deardorff* (pp. 43–51). World Scientific Studies.

Krugman, P., Melitz, M., & Obstfeld, M. (2012). *International economics: Theory and policy, global edition*. Pearson.

Kuorikoski, J., Lehtinen, A., & Marchionni, C. (2010). Economic modelling as robustness analysis. *The British Journal for the Philosophy of Science*, *61*(3), 541–567. https://doi.org/10.1093/bjps/axp049

Ladyman, J. (2019). Structural realism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2019). Metaphysics Research Lab, Stanford University.

Ladyman, J., & Ross, D. (n.d.). *Everything must go: Metaphysics naturalized*. Oxford University Press.

Landry, E. (2007). Shared structure need not be shared set-structure. *Synthese*, *158*(1), 1–17. https://doi.org/10.1007/s11229-006-9047-7

Leamer, E. (1995). *The Heckscher-Ohlin model in theory and practice*. International Economics Section, Departement of Economics Princeton University.

Leamer, E. (2012). *The craft of economics: Lessons from the heckscher-ohlin framework*. MIT Press.

Leamer, E., & Levinsohn, J. (1995). International trade theory: The evidence. *Handbook of International Economics*, *3*, 1339–1394. https://doi.org/10.1016/S1573-4404(05)80006-1

Leamer, E., & Stern, R. (2006). *Quantitative international economics*. Routledge.

Lehtinen, A. (2016). Allocating confirmation with derivational robustness. *Philosophical Studies*, *173*(9), 2487–2509. https://doi.org/10.1007/s11098-016-0625-3

Lehtinen, A. (2018). Derivational robustness and indirect confirmation. *Erkenntnis*, *83*(3), 539–576. https://doi.org/10.1007/s10670-017-9902-6

Leontief, W. (1953). Domestic production and foreign trade: The American capital position re-examined. *Proceedings of the American Philosophical Society*, *97*(4), 332–349.

Lisciandra, C. (2017). Robustness analysis and tractability in modeling. *European Journal for Philosophy of Science*, *7*(1), 79–95. https://doi.org/10.1007/s13194-016-0146-0

Mäki, U.. (2006). On the method of isolation in economics. In J. Davis & J.B. Davis (Eds.), *Recent developments in economic methodology* (Vol. 3, pp. 3–37). Edward Edgar.

Mäki, U. (2009). MISSing the world: Models as isolations and credible surrogate systems. *Erkenntnis*, *70*(1), 29–43. https://doi.org/10.1007/s10670-008-9135-9

Massimi, M. (2018). Perspectival modeling. *Philosophy of Science*, *85*(3), 335–359. https://doi.org/10.1086/697745

McCallum, J. (1995). National borders matter: Canada-US regional trade patterns. *The American Economic Review*, *85*(3), 615–623.

Morgan, M., & Morrison, M. (1999). *Models as mediators*. Cambridge University Press.

Morrison, M. (1999). Moels as autonomous agents. In M. Morgan & M. Morrison (Eds.), *Models as mediators* (pp. 38–65). Oxford University Press.

Morrison, M. (2011). One phenomenon, many models: Inconsistency and complementarity. *Studies in History and Philosophy of Science Part A*, *42*(2), 342–351. https://doi.org/10.1016/j.shpsa.2010.11.042

Morrison, M. (2015). *Reconstructing reality: Models, mathematics, and simulations*. Oxford University Press.

Parker, W. (2010a). Predicting weather and climate: Uncertainty, ensembles and probability. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, *41*(3), 263–272. https://doi.org/10.1016/j.shpsb.2010.07.006

Parker, W. (2010b). Whose probabilities? Predicting climate change with ensembles of models. *Philosophy of Science*, *77*(5), 985–997. https://doi.org/10.1086/656815

Rodrik, D. (2017). *Economics rules: Why economics works, when it fails, and how to tell the differences*. Oxford University Press.

Ross, D. (2008). Ontic structural realism and economics. *Philosophy of Science*, *75*(5), 732–743. https://doi.org/10.1086/594518

Ross, D. (2014). Philosophy of macroeconomics and economic policy. In *Oxford handbooks online*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199935314.013.47

The Royal Swedish Academy of Science (2008). *The prize in economic sciences* [Press release: Trade and geography – economies of scale, differentiated products and transport costs].

Suppe, F. (1989). *The semantic conception of theories and scientific realism*. University of Illinois Press.

Suppes, P. (1960). A comparison of the meaning and uses of models in mathematics and the empirical sciences. *Synthese*, *12*, 287–301. https://doi.org/10.1007/BF00485107

Suppes, P. (2002). *Representation and invariance of scientific structures*. CSLI Publications (distributed by Chicago University Press).

Teller, P. (2001). Twilight of the perfect model model. *Erkenntnis*, *55*(3), 393–415. https://doi.org/10.1023/A:1013349314515

Tinbergen, J. (1962). *Shaping the world economy*. The Twentieth Century Fund.

Van Bergeijk, P. A., & Brakman, S. (2010). *The gravity model in international trade: Advances and applications*. Cambridge University Press.

Van Fraassen, B. C. (1980). *The scientific image*. Oxford University Press.

Weisberg, M. (2006). Robustness analysis. *Philosophy of Science*, *73*(5), 730–742. https://doi.org/10.1086/518628

Weisberg, M. (2012). *Simulation and similarity: Using models to understand the world*. Oxford University Press.

Weisberg, M., & Reisman, K. (2008). The robust Volterra principle. *Philosophy of Science*, *75*(1), 106–131. https://doi.org/10.1086/588395

Ylikoski, P., & Kuorikoski, J. (2010). Dissecting explanatory power. *Philosophical Studies*, *148*(2), 201–219. https://doi.org/10.1007/s11098-008-9324-z

Yotov, Y. V., Piermartini, R., Monteiro, J. A., & Larch, M. (2016). *An advanced guide to trade policy analysis: The structural gravity model*. World Trade Organization.