



The Evidence for Free Trade and Its Background Assumptions: How Well-Established Causal Generalisations Can Be Useless for Policy

Luis Mireles-Flores

To cite this article: Luis Mireles-Flores (2022) The Evidence for Free Trade and Its Background Assumptions: How Well-Established Causal Generalisations Can Be Useless for Policy, Review of Political Economy, 34:3, 534-563, DOI: [10.1080/09538259.2021.1912484](https://doi.org/10.1080/09538259.2021.1912484)

To link to this article: <https://doi.org/10.1080/09538259.2021.1912484>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 29 Apr 2021.



[Submit your article to this journal](#)



Article views: 2751



[View related articles](#)



[View Crossmark data](#)

The Evidence for Free Trade and Its Background Assumptions: How Well-Established Causal Generalisations Can Be Useless for Policy

Luis Mireles-Flores 

University of Helsinki, Helsinki, Finland

ABSTRACT

In this article, I offer a methodological analysis of the empirical research on the causal effects of trade liberalisation, and assess whether such studies can be of any use for guiding policy prescriptions in real-world economies. The analysis focuses on the mainstream economic research that has been used to support arguments in favour of trade liberalisation during the last decades. Even though there are empirical results that could be taken as valid evidence for a causal connection between free trade and economic gains, none of the existing evidence licences trustworthy inferences about the policy effectiveness of trade liberalisation reforms in real-world cases. There are three aspects of the empirical literature that make it highly problematic for making reliable policy inferences: (a) the criteria used to define the notion of 'free trade', (b) the background assumptions embedded in the econometric techniques used for estimating causal effects, and (c) the widespread desire among academic economists to attain scientific results in terms of universally valid generalisations. The analysis exposes a worrisome mismatch between, on the one hand, the research aims and outcomes of scientific economics and, on the other, the kind of evidence that would be useful for guiding actual policy deliberations.

ARTICLE HISTORY

Received 11 February 2020

Accepted 5 January 2021

KEYWORDS



Economic methodology; free trade; empirical evidence; causal inference; policy relevance; science-policy interactions

JELCodes

B41; C18; F13; O24

1. Introduction

Whether free trade causes economic gains or not is one of the oldest unsettled questions of political economy (see Mill [1844] 1874). In the public policy arena, international trade is a recurring topic of heated controversy. Among academics, endorsements of free trade (e.g. Feenstra 2006; Lal 2006; Irwin 2009) and criticisms (e.g. Brown 2006; Chang 2008; Fletcher 2010) come from a wide range of theoretical, political, and ideological stand-points. Driving this longstanding debate is an underlying intuition that the systematic results of scientific research on the nature and consequences of free trade can be straightforwardly exploited to make reliable policy prescriptions. This intuition is misleading.

CONTACT Luis Mireles-Flores  luis.mireles-flores@helsinki.fi  TINT, Practical Philosophy, Faculty of Social Sciences, University of Helsinki, Helsinki, Finland

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Before the 1980s, international economists concerned with trade liberalisation focused on debating the theoretical validity of economic insights from classical political economy and neo-classical economics. Researchers typically used formalised characterisations of notions like comparative advantage, competitive markets, monopolistic competition, trade barriers, and increasing returns to scale, in order to demonstrate either the beneficial effects of free trade (e.g. Dixit and Norman 1980; Krugman 1993; Neary 2003) or the conceptual flaws in neoclassical accounts (e.g. Steedman 1979; Kitson and Michie 1995; Daly 1996; Driskill 2012). However, a final consensus has never been reached as to which theory (or theories) offers the correct view. The lack of a decisive account at the theoretical level, in favour of or against free trade, generated a strong wave of applied empirical studies—mainly during the 1980s and 1990s—with the aim of testing the available theoretical proposals as well as some of the posited effects of trade liberalisation.

By the turn of the century, many economists, policy makers, and other end-users of economics had become convinced that the results of this swelling empirical research finally provided conclusive evidence for the claim ‘free trade causes economic gains’. This claim is a causal generalisation or, as I will refer to it in this article, a *causal-efficacy hypothesis* about free trade. Simultaneously, the same corpus of empirical results was taken as evidence for the validity of a closely connected, yet distinct, claim: ‘to enjoy economic gains, a country should reduce its barriers to international trade’. This is a policy prescription, or what I will call a *policy-effectiveness hypothesis* about free trade.

In order to agree with the policy-effectiveness hypothesis exclusively on the grounds of empirical evidence in favour of the causal-efficacy hypothesis, one has to accept that the validity of a scientific generalisation of the form ‘A causes B’ can entail or warrant the validity of a policy prescription of the form ‘If you want B, you should do A’. However, as I will argue, *the validity of the causal-efficacy hypothesis does not at all entail or warrant the validity of the policy-effectiveness hypothesis*. I develop the argument for this point by analysing in detail some of the most influential empirical studies on the benefits of free trade from the last four decades. Can the available empirical evidence in favour of free trade be used as a reliable guideline for economic policy making? My conclusion is that the evidence is neither decisive nor even adequate to reliably support policy recommendations in real-world situations. Nonetheless, there are noteworthy lessons to be learned from looking at the details of why this is the case.

The empirical literature analysed in this article is confined to what can be labelled ‘mainstream empirical research’ in favour of free trade, which in turn is closely aligned with the neo-classical economic approach. Thus, the literature I review is not representative of all existing theoretical perspectives on international economics. The selection of this strand is based on the following considerations: (a) These articles are the most highly cited and influential among economists and other social scientists in apologetics for free trade; (b) these studies enjoy a high level of policy impact inasmuch as they have typically been used by international organisations as the evidential base for outlooks and reports regarding trade policy advice (e.g. IMF 1997; OECD 1998; Georgiadis and Gräß 2013); (c) the ultimate goal of my analysis is to bring to the forefront a worrisome mismatch between, on the one hand, the research aims and outcomes of this kind of scientific economics and, on the other hand, the kind of evidence that would in fact be useful for guiding policy deliberations. This problematic mismatch is, I believe, rather endemic and chronic in policy-oriented economic research. However, in the space of one article, I

can only hope to properly analyse one concrete instantiation of it, hence the focus on the case of the empirical research on free trade.

As is well known, heterodox economists have provided substantial criticisms of the neo-classical approach to international trade, concerning both its theoretical foundations and its empirical claims in favour of free trade (e.g. Reinert 2007; Chang 2008; Ocampo, Rada, and Taylor 2009; Taylor 2010). This article takes a look, from a philosophy of science perspective, at a methodological dimension underlying the debate on the benefits of free trade. In particular, I dig into the empirical literature to investigate the various kinds of methodological background assumptions that, in spite of enjoying some epistemic justification according to the discipline's scientific standards, generate unavoidable limitations to the policy relevance of the results put forward as evidence in favour of free trade. Critiques of mainstream arguments for free trade rarely focus on this underlying dimension, so my analysis can be considered as complementary to the existing accounts. Moreover, similarly focused methodological assessments could also be made about any other economic approach (neo-classical or not) that claims to derive reliable policy prescriptions directly from its theoretical or empirical framework.

In Section 2, I briefly portray the kind of policy aspirations underlying most scientific research in international trade economics. In Section 3, I analyse the notion of 'trade liberalisation' and the main econometric techniques used in mainstream empirical literature, and then show how different kinds of methodological background assumptions heavily shape the causal meaning of the research results. In Section 4, I point out and discuss three aspects of the empirical studies that make the results highly problematic for inferring the effectiveness of policy prescriptions. In Section 5, I elaborate on the kind of evidence that would be more in line with the contextual requirements of the policy-making process in real-world economies.

2. The Policy Relevance Aspirations of International Trade Economics

Even if the ultimate aim of economic theory is better policy, one does not best serve that aim by trying to make every journal article into a policy proposal. The immediate policy implications of a new idea are in the end less important than its intellectual contribution. (Krugman 1993, p. 366)

Contrary to Paul Krugman's opinion, economics articles often devote several lines, or an entire section, to highlighting what the immediate policy implications of the main results are supposed to be. According to Milberg (1996), the specialised literature on international trade is particularly fond of this practice and tends to present, motivate, and interpret most questions and answers as having immediate and clear-cut policy implications.¹ For instance, supporters of the so-called new international economics have assumed from the outset that new trade theory explains why reductions to trade barriers generate economic gains (in their models), and therefore tariff reductions are automatically put forward as good policy prescriptions (in the real world). In this theoretical approach, the hypothesis that 'trade liberalisation causes economic gains' is not a theoretical finding, but a working presupposition. In the words of Robert Feenstra:

¹Milberg (1996) identifies a number of specific rhetorical devices that professional economists use to make international trade articles look more policy relevant than they may actually be.

[T]he models of economies of scale and monopolistic competition *were conceived with a very practical application in mind, namely, the gains that would result from large-scale tariff reductions*. Whether from multilateral tariff reductions under the WTO, or bilateral tariff reductions under regional trade agreements, these models predicted gains from trade over and above the gains from specialization in conventional models. (Feenstra 2006, pp. 617–618; emphasis added)

Similarly, the empirical studies on the benefits of free trade that proliferated during the 1980s and 1990s had a clear policy-relevance motivation as well. In an influential review of the literature, Rodríguez and Rodrik (2001) describe what they take as the main question driving applied international trade research:

Do countries with lower policy-induced barriers to international trade grow faster, once other relevant country characteristics are controlled for? We take this to be *the central question of policy relevance* in this area. To the extent that the empirical literature demonstrates a positive causal link from openness to growth, *the main operational implication is that governments should dismantle their barriers to trade*. (Rodríguez and Rodrik 2001, p. 264; emphasis added)

This makes explicit what seems to be a tacit consensus in applied international economics, i.e. that the causal hypothesis ‘free trade causes economic growth’ is crucial and deserves scientific investigation because of the significance of its expected policy implications. The underlying intuition—which in the latter quote is rather explicit—is that, to the extent that the posited causal relation between free trade and economic growth (*a causal-efficacy hypothesis*) is empirically validated, then the policy prescription that ‘governments should dismantle their barriers to trade’ (*a policy-effectiveness hypothesis*) will automatically follow. As mentioned in the introduction, this is precisely the widespread intuition that I claim is misleading.

A strong emphasis on turning scientific results into policy relevant prescriptions has led some economists to mistakenly consider the potential or conditional policy implications of scientific research as automatically applicable in actual situations. As a consequence, there is a great risk, on the one hand, of researchers believing that the scientific knowledge they produce is much more suited and relevant to guiding policy than it actually is, and on the other hand, of the end-users of science (i.e. policy makers, governments, and the general public) believing that the scientific knowledge they receive is unconditionally reliable and useful for achieving specific policy goals.

3. Empirical Evidence of the Benefits of Free Trade

Empirical studies on the economics of free trade are customarily of two broad types. One type aims at testing the validity of a given theory. For instance, new trade theory postulates that monopolistic competition models can be used to formally derive the following argument: increasing the level of foreign trade will generate (a) efficient cost reductions, (b) a wider variety of available goods in the domestic market, and (c) decreases in the price levels within the trading countries (see Feenstra 2006). Accordingly, there are studies designed primarily as empirical tests of these theoretical implications.

The other type of empirical research focuses on evaluating the effects of trade-liberalisation reforms. The primary goal of these studies is not to test underlying theories, but rather to estimate the causal influence of trade liberalisation on other economic variables,

using macroeconomic statistical data. Even if these studies are presented as mainly empirical, they are still heavily theory-based in fundamental ways (clarified below). It is this second type of study that is most commonly cited as evidence in favour of trade liberalisation and the focus of my methodological analysis.

For the most part, the research on the benefits of free trade consists of econometric estimations of the causal efficacy of a stylised variable standing for trade liberalisation. The hypotheses under evaluation are typically of the generic form:

‘(For P), TL causes Y ’

where P is the statistical population of countries from which the dataset has been collected; TL is a variable standing for trade liberalisation; and Y stands for any of the usual indicators of national economic performance, such as growth per capita, investment, income equality, price levels, and so on, as measured by conventional international standards.

As concrete illustrations, I present here a non-exhaustive list of hypotheses about the economic gains (Y) of trade liberalisation (TL), which have been empirically tested using the concepts, causal criteria, and inferential methods that will be examined in this article:

- (1) TL causes increases in economic growth (GDP/capita).²
- (2) TL causes increases in investment (INV).³
- (3) TL causes increases in the proportion of trade with foreign countries.⁴
- (4) TL causes reductions in the prices of goods, given increasing returns to scale.⁵
- (5) TL causes increases in the variety of goods available to consumers.⁶
- (6) TL causes increases in economic competition, which in turn causes the self-selection (and survival) of the most efficient firms.⁷
- (7) TL causes convergence of wages, which leads to reductions in income inequalities.⁸
- (8) TL causes reductions in unemployment.⁹
- (9) TL causes transfers of foreign technologies.¹⁰

To the extent to which the causal-efficacy hypotheses listed above are taken as empirically established according to the relevant scientific standards in the field, they can be taken as evidence in favour of the broad claim that trade liberalisation causes economic gains. But what exactly is one justified in inferring (and what not) from the results of these empirical studies?

In the rest of Section 3, I will highlight two features of the empirical research on free trade that are crucial to the policy relevance of its results. First, in 3.1, I show how the

²See Edwards 1992; Dollar 1992; Sachs and Warner 1995; Greenaway, Morgan, and Wright 1997; Ades and Glaeser 1999; Wacziarg and Horn Welch 2008.

³Measured in terms of INV to GDP ratio (INV/GDP), and typically via foreign direct investment (FDI) as a mediating variable; see Baldwin and Seghezza 1996; Wacziarg and Horn Welch 2008.

⁴Measured in terms of the so-called ‘openness ratio’; see Wacziarg and Horn Welch 2008.

⁵See Harris 1984; Smith and Venables 1988; Badinger 2007.

⁶See Feenstra 1994; Hummels and Klenow 2005; Broda and Weinstein 2006.

⁷See Coe, Helpman, and Hoffmaister 1997; Bernard et al. 2003; Eaton, Kortum, and Kramarz 2004; Trefler 2004; Feenstra and Kee 2008.

⁸See Ben-David 1993; Frankel and Romer 1999; Winters, McCulloch, and McKay 2004.

⁹See Krueger 1983; Milner and Wright 1998; Falvey 1999; Dutt, Mitra, and Ranjan 2009.

¹⁰Mainly in the form of foreign capital equipment and technological research; see Keller 2004; Madsen 2007.

methodology employed for testing the causal hypotheses interprets the notion of ‘trade liberalisation’ (TL), that is, what the variable TL means in these studies. Then, in 3.2, I examine the notion of causation that is implicit in the methodology employed to test the causal hypotheses, or in other words, what exactly a ‘causal effect’ means in these studies. Different interpretations of trade liberalisation as well as different notions of causation would enable or constrain the inferences that can be drawn about the effects of policy prescriptions from scientific results.

3.1. The Meaning of ‘Trade Liberalisation’

The first thing that comes to mind when one thinks of trade liberalisation is the elimination of taxes on international exchange. This is somewhat correct, since trade liberalisation reforms almost always include the reduction of commercial tariffs. This is also the way in which typical theoretical analyses study the conceptual effects of trade liberalisation, for instance, by calculating the burden of international tariffs on different economic sectors. But what exactly is the meaning of ‘trade liberalisation’ in empirical research? How is it measured? What kind of change in a real economy corresponds to a change in the variable ‘trade liberalisation’?

In empirical studies, international economists define a variable TL—often referred to as *openness*—that measures liberalisation levels in terms of the amount of trade restrictions in a particular country for a certain year. Given that many countries have initiated clear-cut liberalisation reforms during the last few decades, researchers have focused on characterising two aspects of liberalisation. These aspects are: (1) a set of *openness criteria* for what counts as an open and as a closed economy, which allow for straightforward comparisons of several economic indicators between open and closed countries; and (2) the precise *liberalisation dates* for all countries that have launched liberalisation reforms, which allow for comparisons of the economic trends and development within each particular country before and after liberalisation reforms were implemented.

The most well-known, and still widely used, *openness criteria* were developed by Sachs and Warner (1995). Using the ‘international comparisons’ dataset by Summers and Heston (1991), Sachs and Warner constructed a dummy variable that takes the value 0 for a country in a particular year if it is closed, and 1 if it is open. A country is considered to have a closed economy in a particular year if at least one of the following five features is true for that country in that year:

- (1) Average tariff rate level of 40 percent or more.
- (2) Non-tariff barriers covering 40 percent or more of all trade.
- (3) Black-market exchange rate at least 20 percent lower than the official exchange rate.
- (4) State monopoly on exports.
- (5) Socialist (centrally-planned) economic system.

Establishing precise *liberalisation dates* is less straightforward, and so various procedures have been tried. A relatively simple method is to use an ex-ante approach and consider, for instance, the ‘statements of intent’ made by countries when a World Bank Structural Adjustment Loan (SAL) is granted. The date on which the loan begins can then be taken

as the starting date of a liberalisation reform (e.g. Harrigan and Mosley 1991; World Bank 1993; Greenaway, Morgan, and Wright 2002).

The alternative is to use an ex-post approach in which a set of countries' economic characteristics are assessed during a certain period to detect significant changes in their openness conditions. For example, Dean, Desai, and Riedel (1994) inferred liberalisation dates for 32 countries from the 1980s to the beginning of the 1990s by analysing in detail their socioeconomic history, focusing on four variables: changes in average tariffs, changes in quotas, export taxes, and foreign exchange restrictions (Dean, Desai, and Riedel 1994, pp. 11–14). Similarly, in the study by Sachs and Warner (1995) already mentioned, the authors inferred liberalisation dates for 111 countries using an ex-post approach and their own—as described above—openness criteria.

An additional variable typically included in empirical studies is the *openness ratio*: a measure of the volume of a country's foreign trade relative to its national product, calculated as a ratio of the imports plus exports to the GDP (in a particular year). It is important to note the contrast with the previously discussed variable for TL, the *openness variable*, which measures the amount of trade barriers a country has in place, whereas the *openness ratio* refers to the proportion of foreign trade a country actually experiences relative to its GDP.¹¹

3.2. Empirical Methods and the Implicit Meanings of Causal Effects

The two most common approaches for estimating the effects of trade liberalisation are cross-sectional and time-series analyses. Cross-sectional analysis—called cross-country analysis when units are countries—has been widely used in international trade and economic growth research, especially during the 1980s and 1990s. In this type of study, a set of explanatory variables (including TL) and a dependent variable (Y) are incorporated into the specification of a regression equation and the influence of each variable on Y is estimated at a particular point in time for all sampled countries. Using matching techniques, it is possible to compare countries that have experienced a trade liberalisation reform (TL = 1) with countries that have not (TL = 0). Assuming that the regression equation includes variables for all relevant causal factors that have a significant effect on Y (apart from the effect of TL), then the result is an estimate of an *average causal effect* of TL on Y, inferred from cross-sectional comparisons between liberalised and non-liberalised countries during the same period.

By contrast, time series analysis offers estimates of the causal effect of trade liberalisation (TL) within one and the same country across a number of consecutive points in time. Using data on the liberalisation dates, it is possible to evaluate trends within countries and significant jumps or breaks in the evolution of the economic indicators as a consequence of trade liberalisation reforms. TL is set to 0 for all years before the liberalisation date of a particular country, and to 1 afterwards. Again, the result is an estimate of an *average causal effect* of TL on Y, but in this case it is inferred from differential comparisons of the same country's variations in Y from one year to the next across a number of consecutive years.

¹¹Changes in liberalisation levels as reflected in the *openness variable* need not coincide with changes in the *openness ratio*, for example, a reduction in the level of trade tariffs (a change in the openness variable) may or may not result in an increase of the proportion of foreign trade relative to GDP (the openness ratio). Similarly, the openness ratio may very well change for reasons other than changes in the level of foreign-trade barriers.

Empirical studies can combine both cross-country and intertemporal analyses whenever panel datasets are available. A panel dataset encompasses information on a relatively large number of socioeconomic characteristics of different countries in different years. By applying the openness criteria to distinguish *liberalised* and *non-liberalised* countries, and then using information from a panel dataset, it is possible to obtain estimates for different groupings of countries in the same study. For instance, one could use a sub-sample exclusively composed of countries that have experienced trade reforms to make before-and-after comparisons of the effects of TL, and one could also use a sample including both liberalised and non-liberalised countries to make comparisons between the different trends followed by open and closed economies (e.g. Greenaway, Morgan, and Wright 1997).

Both types of study are meant to assess whether TL has any causal influence over Y. But is it really the case that these two types of method are actually testing the same thing?

The fact that causation is a pluralist notion and that different causal connotations of empirical results are dependent on the methods of causal inference employed to establish them has long been recognised and discussed in philosophy of science (see Hitchcock 2003; Cartwright 2007). In the case at hand, it is clear that an *average causal effect* will mean different things depending on which empirical method (cross-country or within-country analysis) is used to establish it. Using different empirical techniques to infer causality implies that the estimates are significantly dependent on the distinct methodological assumptions that are essential to each method (e.g. assumptions about country homogeneity, potential confounders, the uniformity of background characteristics, and so on). Thus, it is important to examine further what these methodological assumptions consist of and why they have to be made in the first place.

3.2.1. *The Role of Background Assumptions in Cross-Country Analysis*

A typical scientific approach to the evaluation of causal hypotheses like ‘TL causes Y’ is to investigate the existence and workings of the causal relation in isolation (see Mill [1843] 1874, 3.8; Mäki 1992; Reiss 2008). The common feature among different econometric methods of causal inference is precisely that they all aim at the best possible way of controlling the potential confounders that could have an effect on Y apart from the posited cause TL. They all differ, however, in the specific means employed to achieve such ideal isolation or shielding.

In cross-country empirical studies, the main way to control for potential confounders consists in trying to include all (and nothing else but) the relevant variables in the regression equation, i.e. getting the correct *model specification*. A common procedure to decide which variables are to be included has been to look at what economic theory has to say. In the case of TL, typically, the starting point is a ‘core’ new-growth-theory model ‘of the type which has now become standard’ in empirical macroeconomics (Greenaway, Morgan, and Wright 2002, p. 234). These models amount to a set of theoretical background assumptions, which are rarely explicitly discussed, yet always present.¹² For

¹²In the mainstream empirical research on international trade, ‘standard specification’ refers to one in line with the models proposed by Romer (1990) and Barro (1991), and with the specification-search results of Levine and Renelt (1992). Why these particular models are the undisputed theoretical basis for growth cross-country regressions, rather than any other, is an important methodological question that should be further investigated. For a detailed

instance, in the highly-cited article by Sachs and Warner (1995),¹³ the authors use Barro's (1991) growth specification as a baseline for their first regression and, after a few variations, they settle on using the following variables:

Dependent variable:

Y: Real GDP per capita

Explanatory variables:

X₁: Sachs-Warner openness dummy variable (TL)

X₂: Ratio of real gross domestic investment to real GDP (INV/GDP)

X₃: Population density

X₄: Secondary school enrolment rate

X₅: Primary school enrolment rate

X₆: Ratio of government consumption to GDP

X₇: Extreme political repression and unrest

X₈: Number of revolutions per year

X₉: Number of assassinations per capita per year

Controlling for this set of variables and using data on 111 countries from 1970 to 1990, Sachs and Warner's cross-country analysis generated an estimate of a positive causal effect of 2.44 from TL on GDP per capita. The magnitude of this estimate means that on average countries classified as open have experienced an economic growth of 2.44 percentage points higher than countries classified as closed.¹⁴

In 2001, Rodríguez and Rodrik published a very influential critical review of the state of the empirical research on the economic effects of trade liberalisation.¹⁵ They reviewed in detail some of the most important studies at the time of the mainstream research (i.e. Dollar 1992; Ben-David 1993; Sachs and Warner 1995; Edwards 1998; Frankel and Romer 1999; and more briefly, Lee 1993; Harrison 1996; Wacziarg 2001), and discussed thoroughly a number of methodological issues in them. According to the authors, the relationship between TL and economic growth had not been accurately assessed in these studies and remained 'far from being settled on empirical grounds' (Rodríguez and Rodrik 2001, p. 266).

Their two main criticisms were directed at the Sachs-Warner openness criteria and at the use of cross-country analysis. On the first issue, they concluded that the Sachs-Warner openness indicator 'yields an upward-biased estimate of the effects of trade restrictions' (Rodríguez and Rodrik 2001, p. 282), and that ultimately it is 'so correlated with plausible groupings of alternative explanatory variables [...] that it is risky to draw strong inferences about the effect of openness on growth based on its coefficient in a growth regression' (Rodríguez and Rodrik 2001, p. 292). Regarding the reliance on cross-country regressions, they argued that the static estimates in cross-sectional analysis could mask dynamic variation in causally relevant country characteristics, which in

discussion of different specification-search methods to find robust determinants of growth for cross-country regressions, see Hoover and Perez 2004.

¹³6898 citations in Google Scholar (by February 2020).

¹⁴The same control variables for the regression specification, and thus the same theoretical background assumptions, have been used with very few variations in subsequent empirical studies, see, e.g., Dollar 1992; Ben-David 1993; and Edwards 1998. Greenaway, Morgan, and Wright 2002 also use the same model specification to test the robustness of TL on GDP per capita to distinct ways of obtaining liberalisation dates. Wacziarg and Horn Welch 2008 replicate Sachs and Warner's result using exactly the same econometric specification but on a different dataset.

¹⁵4973 citations in Google Scholar (by February 2020).

turn could make the validity of the estimates not time-invariant. Overall, they concluded that the results of the empirical research on free trade, at the time, could not be trusted:

For the most part, the strong results in this literature arise either from obvious misspecification or from the use of measures of openness that are proxies for other policy or institutional variables that have an independent detrimental effect on growth. When we do point to the fragility of the coefficients, it is to make the point that the coefficients on the openness indicators are particularly sensitive to controls for these other policy and institutional variables. (Rodríguez and Rodrik 2001, p. 315)

3.2.2. *The Turn to Within-Country Analysis*

As a consequence of Rodríguez and Rodrik's criticisms, subsequent studies reduced their reliance on cross-country estimation in favour of within-country estimation using time series. This alternative technique can be used to analyse the effects of trade reforms over time without relying so much on choosing adequate controls for comparing heterogeneous countries. Panel data analyses in which cross-country estimations were complemented by within-country estimation techniques subsequently started to appear. In particular, adaptations of econometric methods following the logic of design-based econometrics and of the potential outcomes framework, such as difference-in-differences and fixed effects analyses, have since become more popular in the literature on international trade.

After Rodríguez and Rodrik's (2001) critique, one of the most influential empirical studies has been Romain Wacziarg and Karen Horn Welch's (2008) 'Trade Liberalization and Growth: New Evidence'.¹⁶ The authors set out to achieve three goals in this article: first, to update the Sachs-Warner openness classification using a more comprehensive database; second, to replicate (and revise) Sachs and Warner's cross-sectional positive results using the updated classification; and it is their 'third and most important goal [...] to exploit the timing of liberalization in a within-country setting to identify the changes in growth, investment, and openness associated with discrete changes in trade policy' (Wacziarg and Horn Welch 2008, p. 189).¹⁷

Using a dataset for 141 countries (from the 1970s until the end of the 1990s) with corrected liberalisation dates, Wacziarg and Horn Welch replicated the cross-sectional analysis and tested the robustness of the estimates for three different decades. They found that the Sachs-Warner results barely held for the 1970s and 1980s, and failed to hold for the 1990s. Their explanation was that relevant country characteristics do indeed vary through time, thus confirming Rodríguez and Rodrik's concerns about the fragility of cross-country estimates. As Wacziarg and Horn Welch put it, their 'results suggest that the Sachs-Warner cross-sectional findings are highly sensitive to the decade under consideration and that the updated openness indicator can no longer effectively distinguish fast-growing from slow-growing countries' (Wacziarg and Horn Welch 2008, p. 197).

¹⁶1883 citations in Google Scholar (by February 2020).

¹⁷These variables correspond to the first three hypotheses listed at the beginning of this section, i.e., the effects of TL on growth, investment, and the openness ratio. The three hypotheses are commonly tested together in the same empirical studies, since investment and the proportion of foreign trade are taken as mediating variables in the causal paths connecting TL to growth.

The authors then used the same dataset to estimate within-country effects of trade liberalisation (TL) on GDP per capita, investment, and volumes of trade relative to GDP (openness ratio). Using equations that amount to difference-in-differences regressions, they found that, in contrast to the rather weak and unreliable cross-country results, ‘the results based on within country variation suggest that over time the effects of increased policy openness within countries are positive, economically large, and statistically significant’ (Wacziarg and Horn Welch 2008, p. 189).

The specification of the regressions and the methodological assumptions in the analysis of within-country effects are entirely different from those used in cross-country analysis in several technical respects (see Wooldridge 2010, ch. 6.5). The main distinction to be noticed here is that, instead of having to explicitly include explanatory variables to control for confounders, the within-country methodology has no need to figure out the right set of independent variables. The estimates are obtained entirely from measuring whether variations in the dummy variable TL relate to a significant change in the average differences of Y within each country from one period to the next. This amounts to contrasting the trend of the yearly variation of Y between pre- and post-liberalisation periods for all countries sampled. Since the comparanda are periodic trends, it can be assumed that the relevant contextual differences among countries are controlled for (see Angrist and Pischke 2015, ch. 5).

As an illustration, in order to estimate the within-country effects of TL on GDP per capita, Wacziarg and Horn Welch ran difference-in-differences regressions in log income of the following form:

$$\log y_{it} - \log y_{it-1} = \alpha_i + \beta LIB_{it} + \varepsilon_{it}$$

where y_{it} is GDP per capita in country i at time t , and LIB_{it} (their variable for TL) takes the value of 1 if t is greater than the liberalisation year.¹⁸ Using this equation, their estimate of β for the whole period from 1950 to 1998 was 1.42 percentage points of average difference in growth between liberalised and non-liberalised countries. Moreover, the coefficient showed a gradual increase over time when the analysis was performed for three consecutive periods (1950–70, 1970–90, and 1990–98). In particular, β reached a statistically significant value of 2.55 for the 1990s. As the authors observed, ‘[t]hese results stand in sharp contrast to the cross-sectional results: countries that liberalized in the 1990s experienced a larger postliberalization increase in growth than countries that liberalized in any other decade’ (Wacziarg and Horn Welch 2008, p. 200).

Overall, during the 2000s, researchers took into account the concerns about cross-country studies raised by Rodríguez and Rodrik (2001), and responded with a renewed wave of studies employing more sophisticated within-country econometric techniques. Wacziarg and Horn Welch (2008), an emblematic example of this new wave, explicitly present their results as revised (after considering Rodríguez and Rodrik’s criticisms) and more correct estimates of the positive causal effects of TL. At first glance, this approach could give the impression that empirical researchers were actually trying to avoid being dogmatic, by taking criticisms into account and revising flaws in their methods, so that ultimately their results could be considered well-established

¹⁸The residual terms ε_{it} are modelled so as to include country and time fixed-effects; see Wacziarg and Horn Welch 2008, pp. 199–202.

according to accepted scientific standards. In fact, the switch from one estimation technique to another clearly implies merely a switch from one set of methodological (conceptual and statistical) assumptions to another set.

As discussed above, the assumptions employed and the logic behind the statistical causal inference provide a characterisation of how each method interprets the notion of causality. The resulting average causal effects will be as distinct in meaning as the methodological assumptions used to obtain them differ. This point need not be problematic in itself; but it definitely makes it unclear whether the estimates derived from both techniques are in fact measuring the same thing, and even less clear whether the results are indeed more robust when the two kinds of results support each other (as presentations of both types of estimates together in the same study often suggest).

Given that these methodological background assumptions shape the meaning of causal effects, how do they affect the reliability of the empirical results for warranting policy prescriptions?

4. Discussion: How Scientific Results Become Useless for Policy

Based on the previous analysis, I argue here that the criteria used to define the notion of TL and the particular econometric methods employed in empirical research severely limit the relevance of the established causal effects for policy-making decisions. As depicted in the preceding section, the research results suggest that a change in the variable TL has an average causal effect on the target variable Y, but *only when subject to* conceptual, statistical, and methodological presuppositions.

In what follows, I shall emphasise three aspects of the empirical results that, on the one hand, are a direct consequence of the a priori assumptions required to carry out the research (and thus might be acceptable from a scientific perspective) but, on the other hand, are highly problematic for the use of the scientific outcomes in guiding policy prescriptions (and thus undermine the usefulness of the results from a policy-making perspective).

4.1. Intervening on Multidimensional Variables

The first aspect is the lack of an unambiguous variable to account for changes in the level of trade liberalisation. As mentioned in section 3.1, there have been methodological discussions among international economists on this issue. The main outcome has been a consensus to conceptualise a *compound* or *multidimensional variable* that could account for several characteristics of open and closed economies (e.g. Sachs and Warner 1995). The variety of proposals for how to capture the notion of TL has in fact motivated empirical assessments of the robustness of results to different measurements of the TL variable (e.g. Greenaway, Morgan, and Wright 2002).

From a scientific perspective, as long as any of the proposed measures are reasonable indicators of trade liberalisation and can be obtained from the available databases, then researchers can take them as acceptable characterisations of TL in their empirical evaluations of the causal efficacy of TL. However, from a policy-oriented perspective, there is no straightforward way, in practice, of intervening on a multidimensional variable such as the ones used to account for free trade levels. Which specific changes in the trade

policy of a country should a government implement in order to affect variable TL in the required way to achieve a desired policy outcome?

As an example, the Sachs-Warner criteria for TL, as noted before, includes five dimensions: level of tariffs, non-tariff barriers, black-market exchange rate level, market power of state commercial companies, and level of state interventionism in the economy. When the results obtained in a study (using the Sachs-Warner measure of TL) significantly support the claim that ‘TL causes growth in GDP per capita’, these results can be taken as scientific evidence that trade liberalisation can be causally efficacious on economic growth. Nevertheless, the results would say nothing specific about how distinct combinations of changes in the different dimensions of TL should be implemented in order to reliably induce a desired effect on the level of economic growth. What exactly should a policy maker do to induce GDP per capita growth in a specific country? Should tariffs be reduced, should non-tariff barriers be eliminated, should the state monopoly on exports be dissolved, or a combination of all these? Which of the five dimensions of TL must be affected, by how much, and in what way? More concretely put, are the estimated causal effects invariant with respect to the specific policy design and implementation of a change in TL?¹⁹

The process of designing the most appropriate policy reform for obtaining a specific result in a real-world situation is a different and separate affair from the typical scientific investigation of causal effects. Given the measuring criteria used to account for TL, even the strongest scientific evidence in favour of the existence of a causal effect would not be very informative for policy makers as to how the variable TL could or should be affected in real situations so as to reliably attain a desired policy effect.

4.2. Average Causal Effects and Their Implicit Applicability Conditions

Another problematic aspect for policy purposes follows from the specific characterisation of causation implicit in the methods used in the empirical research on free trade. As shown in section 3.2, any coefficient elicited using cross-country and within-country methods, regardless of how statistically significant and unbiased, ultimately represents a particular type of causal concept which is characterised by a specific set of methodological and statistical a priori assumptions. Thus, the validity of the results is entirely dependent on the validity of these assumptions. Of course, all methods of causal inference involve the postulation of assumptions in order to help identify and isolate manifestations of causal efficacy. However, precisely because such a priori assumptions are always required, the empirical results obtained cannot constitute—and should not be taken as—fully reliable evidence for the effectiveness of actual policy reforms in a real-world economy.

From a scientific perspective, the results constitute valid estimates of an average causal effect of TL on other socio-economic variables, such as GDP per capita or the investment to GDP ratio. If the estimates were obtained after following all relevant quality standards for this kind of econometric study, then the results can be taken as scientifically valid evidence in favour of the hypothesis that changes in the TL variable have an average causal

¹⁹For a similar discussion of the potential ambiguity of policy interventions based on multidimensional variables, but in the context of empirical research on the causes of unemployment, see Claveau and Mireles-Flores 2014, pp. 400–402.

effect on variations of the GDP per capita (or on INV/GDP). The resulting estimate is a causal concept inferred from a large number of counterfactual comparisons of country-level datasets (as depicted in Section 3).

From a policy-making perspective, however, these results say nothing about the validity of the postulated causal influence of TL or its expected effect on target variable Y *in the socioeconomic context of any particular country*. This is the case because the methodology for estimating average causal effects is designed specifically to control for all relevant contextual characteristics and all observable and unobservable potential confounders. By contrast, the evaluation and implementation of policy prescriptions and effective policy reforms require exactly the opposite approach: the careful taking into account of (as many as possible of) the *relevant* and *knowable* contextual features and potential disturbing factors related to a specific target situation.

The econometric techniques employed to control for the influence of known and unknown causal factors—in this case cross-country and within-country methods—have their respective merits and deficiencies depending on the epistemic aim at hand. Cross-country methods have the advantage of allowing researchers to explicitly control for confounding factors (such as education level, social conflicts, institutional framework, and so on), which is useful insofar as there is reliable background knowledge about these factors. But since the resulting estimate essentially conveys a cross-sectional picture of economies at particular instants, if countries experience a high socioeconomic variability over time, then the estimates will be unreliable for making accurate and specific policy inferences. Alternatively, within-country methods have the advantage of allowing researchers to explore the trends that countries experience before and after actual trade liberalisation reforms, as well as the trends followed by countries that have not implemented such reforms. But there is often causally relevant heterogeneity across countries underlying the final average results, and thus again the estimates will be unreliable for making accurate inferences about the expected effects of a trade liberalisation reform in any particular country.

Most of the intrinsic methodological limitations of the estimation techniques are not unknown to expert econometricians and empirical economists. In fact, the merit of many contributions to the field consists of technical improvements intended to deal with such methodological issues, e.g. sensitivity tests, multiple regressions to test for robustness of different approaches and specifications, or case-study approaches to deal with single-unit heterogeneity (see Wooldridge 2010; Best and Wolf 2015). Furthermore, empirical researchers seem to be aware of the fact that conceptual and methodological assumptions in their econometric techniques restrict the inferential potential of their results in different ways.

In fact, studies that support the positive economic effects of trade liberalisation often add qualifications (perhaps too inconspicuously so as to not undermine their main results) about the interpretation and limitations of the results outside the specified dimensions of the scientific study. For example, the existence of concurrent policies in real economies (but that are not measured in the studies) is commonly recognised as a potential constraint to the effectiveness of trade policy in the real world (e.g. Greenaway, Morgan, and Wright 1997, p. 1886; Greenaway, Morgan, and Wright 2002, p. 233; and Wacziarg and Horn Welch 2008, pp. 206–207). In other words, researchers are aware, even if they are not very explicit about it, that their results are valid only under *ceteris paribus* conditions.

As a more concrete example, Wacziarg and Horn Welch (2008)—just after presenting their revised positive results—add in passing the following cautionary remark on the potential heterogeneity of single-unit effects, which can always lie beneath average-based estimates:

[T]he extent to which per capita income growth changed after trade reforms varied widely across countries. While the average effect obtained in the large sample is positive, roughly half of the countries experienced zero or even negative changes in growth following liberalization. [...] generalizations about the factors that may explain these differences are difficult to draw. The institutional environment of countries, the extent of political turmoil, the scope and depth of economic reforms, and the characteristics of concurrent macroeconomic policies all seem to have a role to play, to varying degrees in different countries. *While this article paints a picture that is highly favorable to outward-oriented policy reforms on average, it cautions against one-size-fits-all policies that disregard local circumstances.* (Wacziarg and Horn Welch 2008, pp. 189–190; emphasis added)

Nevertheless, awareness about relevant qualms, qualifications, and precautions in relation to the interpretation of the results is absent in almost all discussions at the policy and decision-making level. In particular, policy-related discussions that specifically refer to the study by Wacziarg and Horn Welch tend to make no reference to the qualms explicitly stated by the authors in the quote above. Consider the popular book *Free Trade under Fire*, in which Douglas Irwin (2009) summarises the outcomes of the existing empirical research in favour of trade liberalisation (directly referring to Wacziarg and Horn Welch's results). Irwin makes no substantial mention of any qualms, limitations, or potential biases related to the econometric average results but instead, he flatly states that 'despite shortcomings in method and measurement, cross-country and within-country studies support the conclusion that economies with more open trade policies tend to perform better than those with more restrictive trade policies' (Irwin 2009, p. 54). This is precisely the type of 'one-size-fits-all' policy conclusion that Wacziarg and Horn Welch warned readers not to make.

Admittedly, qualms and qualifications made by academic researchers might be easy to miss in the midst of reading and trying to become acquainted with a large available corpus of empirical studies. Nonetheless, it is an issue for reasonable concern that, more often than not, these kinds of qualms and their significance for policy-making purposes fail to reach public debates or the policy-making arena. Are economists failing at communicating the relevant qualifications for end-users to reach a correct interpretation of their empirical results? Are decision-makers failing at expressing their real-world policy needs to economists or at demanding clear guidelines for the interpretation and application of scientific results? If economists know and understand the inherent shortcomings of their empirical methods (e.g. those related to cross-sectional variation, time variation, parameter heterogeneity, and so forth), why are these well-understood limitations not promulgated among the end-users of the science just as much as the final results? This is an especially important concern since by ignoring these shortcomings the end-users could easily fall into making dogmatic and potentially dangerous misapplications of the scientific results.²⁰

To sum up, the second problematic aspect of the empirical research is that the specific methodology used to test causal effects strongly shapes a priori the meaning of 'causing'

²⁰This kind of dogmatism and its problematic consequences can be clearly noticed, *mutatis mutandi*, by looking at the policy prescriptions that were derived from the Washington Consensus; see Williamson 2000; Rodrik 2006.

in claims like ‘TL causes Y’. Once a meaning for ‘TL causes Y’ gets fixed under a particular set of background assumptions, a predetermined stipulation of when and how that particular causal relation will obtain in real life situations also gets fixed. Of course, making assumptions is a standard and essential practice in scientific inquiry, both when theorising and when doing empirical testing. The problem is not so much that there are assumptions determining the meaning and applicability conditions of the outcomes of scientific research, but rather that these assumptions are rarely explicitly discussed or sufficiently emphasised by the scientists when they put forward scientific results as policy relevant, and moreover, that they are rarely acknowledged by policy makers when they use scientific results to guide policy deliberations.

4.3. A Longing for Universal Validity

The third aspect to be noted can be labelled the ‘saving-the-generalisation attitude’ in scientific research; this is the conviction that the more general a result can be taken to be, the better. General validity (broadly conceived) is considered a virtue of any scientific outcome in most academic economics. As a consequence, whenever an empirical analysis shows significant exceptions to an otherwise well-established causal generalisation, a priority among researchers tends to be to save the truth of the generalisation by trying to explain the outlying cases away.

Wacziarg and Horn Welch’s article provides a clear example of this attitude. After presenting the empirical results of their study—estimates of positive average causal effects of trade liberalisation—they comment on a subsample of countries that actually show non-significant or negative effects after liberalisation reforms when individually considered (Wacziarg and Horn Welch 2008, pp. 207–212). They elaborate substantially on possible non-economic contextual reasons for these exceptions in the form of brief ‘country case studies’ in their Appendix 3, thereby providing ex-post justifications for why the causal generalisation ‘free trade causes economic gains’ failed to obtain in each of these countries. As will become clearer in the next section, this type of country-specific information could in fact be highly useful for increasing the reliability of policy prescriptions, but in their article such additional contextual evidence was relegated to the appendices and examined with the exclusive aim of saving the hoped-for general validity of the average causal hypothesis under study.

One of the main goals of the detailed methodological analysis provided in Section 3 was precisely to show how and to what extent empirical studies—once their assumptions are taken for granted—constitute evidence in favour of a causal-efficacy hypothesis, which in this case comes in the form of an *average causal effect*. From a purely scientific perspective, there is no problem whatsoever with considering the empirical studies about, say, the positive causal effects of TL on GDP per capita, as valid evidence supporting the causal generalisation ‘trade liberalisation causes economic gains’. However, from a policy-making perspective, the validity of such a causal generalisation does not reveal much about how specific trade liberalisation reforms could actually affect specific countries in their particular situations. Without sufficient additional contextual information about the ways in which relevant disturbing factors and causal backgrounds can affect the intended policy outcomes, decision makers will always be in danger of inferring and implementing ineffective, or plainly wrong, policy prescriptions.

5. Reconsidering the Evidential Needs of the Policy-Making Process

So far, I have made clear how, *from a scientific perspective*, the causal results on the benefits of trade liberalisation in mainstream empirical research can be taken as valid and well-established, but only contingent on a set of background assumptions. I have also argued that, *from a policy-making perspective*, the a priori assumptions implicit in the empirical research limit the reliability of any inference from the empirical causal results to the context of particular cases, for which the assumptions might not hold true at all.

To appreciate the significance of this evidential mismatch between the requirements of science and those of policy-making, consider the following two hypotheses:

h_1 : For P , trade liberalisation causes economic growth.

h_2 : For country u , implementing a trade liberalisation policy will generate economic growth.

Hypothesis h_1 is a representation of what I have called a causal-efficacy hypothesis on the benefits of free trade, while h_2 is an instantiation of what I have called a policy-effectiveness hypothesis. Philosophers and other scholars concerned with the policy relevance of science have tended to assume that these two claims are essentially of the same kind, and that h_1 entails h_2 . Indeed, both hypotheses are causal and apparently connect the same kinds of causal relata, but in fact they are significantly distinct.²¹ Most importantly, it would be a mistake to assume that h_2 necessarily or automatically follows from h_1 . Evidence that supports h_1 could be sound and scientifically valid, and yet not be sufficient or relevant to warrant the truth of h_2 .

As the previous sections show, empirical economists have made reasonable progress in relation to evaluating and supporting claim h_1 . They have defined and agreed upon measurement variables that allow them to characterise the level of trade liberalisation in different countries, and more specifically, to define a dummy variable to demarcate open from closed economies. They have employed different econometric techniques to evaluate cross-sectional differences between open and closed economies, and within-country differences before and after trade liberalisation reforms have been implemented. They have obtained positive and significant estimates of the average causal effects from changes in trade liberalisation on the level of GDP per capita, investment, and a few other important macroeconomic variables. All this may be taken as scientifically valid evidence in favour of hypothesis h_1 . Nevertheless, for the reasons discussed in Section 4, none of the evidence that supports the *causal-efficacy hypothesis* h_1 can automatically be taken as valid, sufficient, or even relevant evidence to support the *policy-effectiveness hypothesis* h_2 .

In contrast to the methods used to establish wide-ranging scientific causal hypotheses, the evaluation of policy recommendations requires adopting a more specific and contextual evidential approach.²² In the remainder of this section, I offer a few suggestions for how scientific evidence could play a more reliable role in the process of evaluating the expected effectiveness of particular policy hypotheses, such as h_2 .

²¹For a historical and methodological account of philosophically significant ways in which economic causal generalisations and policy recommendations differ, see Mireles-Flores 2016.

²²The view endorsed in this section is inspired by Harold Kincaid's (2004) contextualist approach to explanation, and by Julian Reiss's (2015) pragmatist theory of evidence.

5.1. Making Methodological Assumptions Explicit

As discussed in the previous sections, a number of conceptual and methodological assumptions have to be made in order to generate meaningful (significant and unbiased) estimates of causal effects from the available data. The assumptions are required to justify the inferential import of the econometric methods employed. Some assumptions, that can be labelled ‘tractability’ assumptions (see Hindriks 2006), are mainly intended to make statistical inference simpler, neater, and more manageable. These assumptions are supposed to be irrelevant to the validity of the empirical results obtained. There are, however, other assumptions, which can be called ‘substantial’ (see Kuorikoski, Lehtinen, and Marchionni 2010), that are connected much more significantly to the validity of the research results. In this subsection, I will offer a few examples of methodological a priori assumptions in the empirical research on the benefits of free trade that are substantial, and that impose significant limitations on the relevance of the results for any reliable policy prescription.

Consider one of the basic assumptions typically made in econometric techniques following the logic of the potential outcomes framework:

Temporal stability: The value of a single-unit effect is assumed to be independent of when the cause is produced and measured in any particular unit of P (see Holland 1986, p. 948).

When this assumption is made, the response of the same unit to the same cause is supposed to be the same over time. It does not matter when any particular unit is exposed to the posited cause (or to its alternative), the response will always be the same for that same unit. Of course, this might hold for some real-life cases, but not for others. In relation to trade liberalisation, this assumption has to be carefully reconsidered. As mentioned in Section 3, one of the problems with the existing empirical evidence on this topic has been precisely that the effects of TL on economic growth seem to vary depending on the decade in which countries have implemented a TL reform. Thus responses to the same treatment in this case are not time invariant (see Wacziarg and Horn Welch 2008, pp. 202–206). Admittedly, econometricians have developed techniques to analyse and deal with these time variations to some extent when estimating average effects (e.g. Firebaugh, Warner, and Massoglia 2013; Brüderl and Ludwig 2015), but it is not at all clear that the nuances and qualifications involved in such technical developments are ever clearly understood or even considered by policy makers. A better understanding of the specific causes or processes responsible for time variations would certainly be useful for any country-specific policy deliberation.

Similarly, let us consider some of the fundamental assumptions that are intrinsic to the difference-in-differences estimation method. The analysis assumes that when calculating the expected values over the differences in variable Y in all countries studied, among which some liberalised at a certain point (i.e. switched from $TL = 0$ to $TL = 1$) and others did not, the result can be treated *as if it were the outcome of a random assignment* of $TL = 0$ or $TL = 1$. Of course, in the real world, there is absolutely nothing random about a country deciding to implement a trade policy reform or not. So the inferences within the econometric study may be justified by assuming random assignment of the treatment, but there is little discussion or qualification offered about how the posited causal effect would be affected in real-world cases, where essentially no policy is ever

implemented in isolation from the other socioeconomic events occurring in those countries.²³

Another fundamental assumption is what is often called the ‘parallel trend assumption’, which states that the average trend of variation observed in the control group (countries during non-liberalised periods) represents the *counterfactual* average trend for the treatment group (countries after liberalisation) had they received no treatment. Ultimately, the causal effect can be interpreted as a difference between the observed trends of Y in liberalised countries and what the values of Y *would have been* with parallel trends if there had been no liberalisation. But how can this substantial assumption be justified in the case of trade liberalisation? Is it reasonable to assume that the trends of countries with no reform are a good representation of how countries that implemented reforms would have evolved if they had not implemented reforms? Furthermore, the fixed-effects techniques—used in Wacziarg and Horn Welch’s difference-in-differences study, as in many other within-country econometric analyses—explicitly introduce substantial assumptions so that a number of unobservable relevant causal characteristics can be taken as time-invariant during different periods for the same country, thereby making the before-and-after comparisons meaningful.²⁴

Overall, the outstanding improvements that within-country econometric techniques have achieved in recent decades allow for fairly reliable estimates of average causal effects even in the face of country heterogeneity and time variation. These improvements, however, come with the cost that any inference about what would happen in a specific real economy would be highly circumscribed by all the required conceptual and methodological assumptions involved in the estimation methods. Somewhat ironically, the more sophisticated the techniques for evaluating causal effects become, the more the results can be accepted by economists as scientifically valid, but the less informative they become in relation to specific country-level policy needs.

But is it really necessary to make the users of science always aware of all the methodological assumptions implicit in scientific research? Perhaps not always and not all of the assumptions, but, depending on the case at hand, some background assumptions definitely turn out to be crucial to the policy relevance of the scientific results. Let us consider another common assumption in most design-based econometric approaches:

Stable unit treatment value assumption (SUTVA): the value of the causal effect of each unit u_i is assumed to be independent of any changes in the causal exposures to TL of any of the other units in P (see Morgan and Winship 2007, pp. 37–40).

This methodological assumption is essential for the reliability of the causal inferences so far discussed. It makes the causal effect invariant to the specific number of individuals receiving and those not receiving the treatment. If this assumption does not hold, then the estimated causal effect could be quite different in different cases depending on the proportion of members of population P that are exposed to the causal variable. Again, this may be a reasonable assumption when estimating some average causal effects, but it seems particularly problematic for making policy inferences in relation to the expected effects of trade liberalisation.

²³For a discussion of this specific problem—known as ‘policy endogeneity’—see Rodrik 2012.

²⁴For more details on the assumptions related to difference-in-differences methods and fixed-effects analysis, see Firebaugh, Warner, and Massoglia 2013; Brüderl and Ludwig 2015; Angrist and Pischke 2015, ch. 5.

In fact, the effects of trade liberalisation for some Asian countries in the 1960s were positive and high, in part because the majority of countries in the region at that moment had closed economies (James, Naya, and Meier 1989; Young 1996). By contrast, the effects of trade liberalisation for Latin-American countries in the 1980s and 1990s were much more ambiguous, especially for countries that started liberalisation policies after most other economies had already opened up their international trade (Agosin and Ffrench-Davis 1995). Moreover, the difference in the effects of trade liberalisation reforms between these two regions could also be a consequence of the historical intra-regional trading networks that were already in place among some Asian countries, but almost non-existent or underdeveloped among Latin-American countries (Tsunekawa 2019).

Trade liberalisation, almost by definition, seems to be a variable to which the SUTVA cannot really apply, since the specific causal effect of TL on country u_i will virtually always depend on whether other trading countries (or at least those in u_i 's economic region) are also implementing TL reforms or not. Consequently, before implementing any liberalising policy reform in country u_i , it seems imperative to evaluate at least how many other countries have already liberalised their markets and to what extent, what the state of the relevant international trade networks is in the relevant area, and how the new overall trading configuration would affect the intended effects for country u_i .

A few experts on causal inference have argued that making methodological assumptions as explicit as possible is certainly a positive move towards a better and more transparent applied science, since leaving them implicit can often lead to misinterpretation and misapplication (see Holland 1986; Spanos 2015). As I have shown, such a recommendation is definitely important in relation to the use of science as a guide for policy making. Whether the assumptions just discussed (or any other similar methodological assumptions made to facilitate the empirical testing of causal relations) are true or not about real countries determines how much one can trust that a causal result will obtain in concrete cases. If the assumptions do not hold in a particular country (even if the average causal effect is positive for a large population of countries), it is necessary to perform additional tests and gather additional contextual evidence of the relevant factors and conditions that could interfere with the realisation of the intended policy effect in that particular country.

5.2. Evaluating Relevant Disturbing and Contextual Factors

The evidential base required to inform a full evaluation of a policy-effectiveness hypothesis like h_2 would consist of all available pieces of reliable knowledge and information (be it scientific or not) about the relevant disturbing and contextual factors that could potentially affect or interfere with the outcome of the specific policy intervention intended in country u_i . In broad terms, the crucial causal knowledge for policy purposes is not about what evidence best supports the hypothesis 'TL causes Y', but rather about the possible ways in which, and the reasons why, the claim 'TL causes Y' might fail to obtain.

Effective policy making requires taking into consideration all sorts of specific conditions of country u_i (that are known or suspected to be relevant disturbing factors), as well as their potential effects on the intended policy outcome. Disturbing and contextual factors need not refer only to economic factors, but to a wide variety of causal influences

relevant to the intended effect, e.g. institutional, geographical, political, the socioeconomic status of the inhabitants, governance conditions, the composition of the economic sectors, cultural features, relevant historical events, other planned and ongoing socioeconomic policies, other countries' trade policies, and so forth. Obviously, our knowledge of these relevant causal factors is in most cases imperfect and partial, and cannot come exclusively from economics, but from all available sources of reliable knowledge (scientific or not). In contrast to the production of scientific research, the process of policy making is an inherently interdisciplinary endeavour. So how are we to know which disturbing and contextual factors could be relevant to specific policy hypotheses such as h_2 ?

An obvious place to begin the investigation of relevant factors is the list of (observable) variables that were controlled for in the empirical studies using regression analysis in the first place. Assuming there are good reasons to believe the specifications used were correct, then policy makers can take seriously some of the parameters estimated in such equations (or at least the signs of their average changes), and then run further more localised empirical tests concerning the effects that each of the confounding exogenous variables could have on the intended policy outcome. For instance, we can consider the variables (listed in section 3.2) that were included in the standard specification used by Sachs and Warner (1995), by Wacziarg and Horn Welch (2008), and by many others in their empirical studies.²⁵

According to the mainstream empirical studies, all the variables in [Figure 1](#), in addition to TL, have an effect on GDP per capita. This is the reason, of course, why all these confounding effects were explicitly controlled for in the observational studies. Hence, the evaluation of a policy-effectiveness hypothesis about a trade liberalisation reform in a particular country will require the assessment, in the local context of the intended policy, of as many causal factors and relations as possible from those depicted in [Figure 1](#). Moreover, it will be useful to search for any additional available information about the contents and details of the set labelled 'unknown variables Z', which could include such things as institutional framework, cultural aspects, or potential influences from concurrent policy reforms. As previously mentioned, this kind of additional information would most likely come from various disciplines, e.g. sociology, political sciences, environmental sciences, anthropology, history, and so on, and not exclusively from economics. The point here is that a clear and explicit account of what was controlled for during empirical testing can provide much more useful knowledge (than study results on their own) about which potential factors should be further investigated in relation to the implementation of particular policy reforms.

5.3. Making Underlying Theories Explicit

Theoretical preconceptions also play a role as background assumptions in the process of evaluating specific policy hypotheses. In economics, there are many different, and often seemingly contradictory, theories that are meant to account for the underlying causal structure of economic phenomena. According to Rodrik (2015), this plurality of accounts should not be understood as a problematic feature of economics, but rather as richness in terms of available tools, both for explanatory and for policy purposes. Different economic

²⁵See footnote 14 above (in subsection 3.2.1), and the references therein.

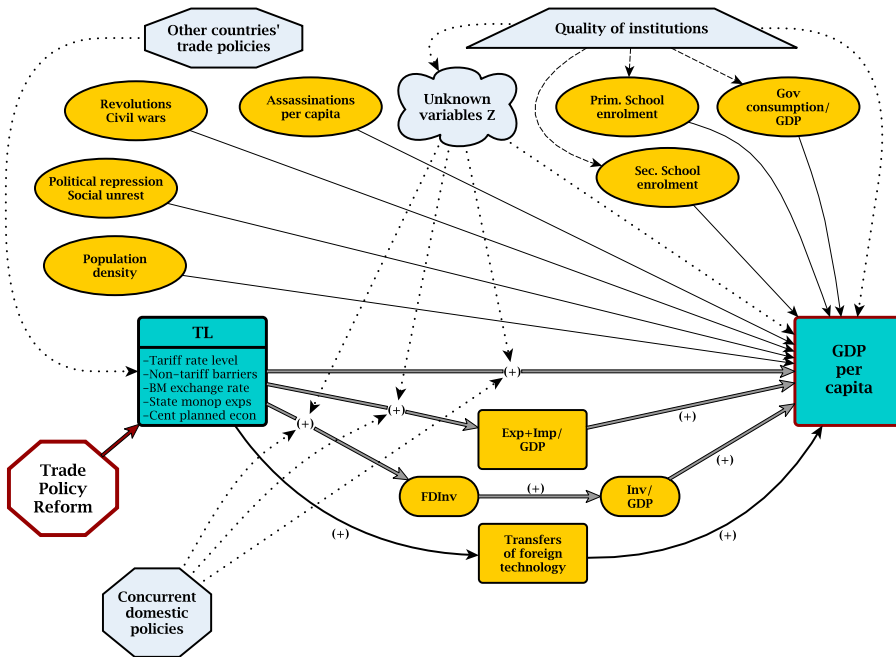


Figure 1. Relevant contextual and disturbing causal factors that are controlled for by the so-called ‘standard specification’ used in growth cross-sectional regressions on the effects of TL.

accounts differ from one another precisely because they are built upon the basis of different assumptions and thus, rather than stating factual contradictions, they state results as conditional, contingent on distinct proposed applicability circumstances (see Rodrik 2015, ch. 3). Of course, the crucial question then is how can one know which theoretical account is the right one to be used in any given situation?²⁶

The minimum scientists should do to address the problem of theory choice is to be as explicit as possible about the theoretical stance implicitly endorsed when engaging in empirical research. For instance, suppose that the monopolistic competition model of new international trade theory (see Krugman 1979, 1980, 1981; Lancaster 1980; Helpman 1981) is the theoretical account favoured by some empirical researchers interested in evaluating whether trade liberalisation causes economic growth. It is very likely that they would also favour the empirical evidence supporting the central tenets of new trade theory (see Feenstra 2006).²⁷ A causal diagram making explicit the connections postulated by this underlying theory can be drawn and used as a complementary guideline for *local* empirical investigations concerned with the implementation of particular trade liberalisation policies.

Figure 2 portrays a number of variables (and their interconnections) as they are postulated and tested by researchers using new trade theory. This causal model can be taken as describing the local applicability conditions of the tenets of the theory. Whether the

²⁶Rodrik offers his own answer to this issue in the form of what he and his colleagues call ‘growth diagnostics’; see Hausmann, Rodrik, and Velasco 2008; and Rodrik 2010.

²⁷See also the footnotes with the references relevant to these tenets at the beginning of section 3.

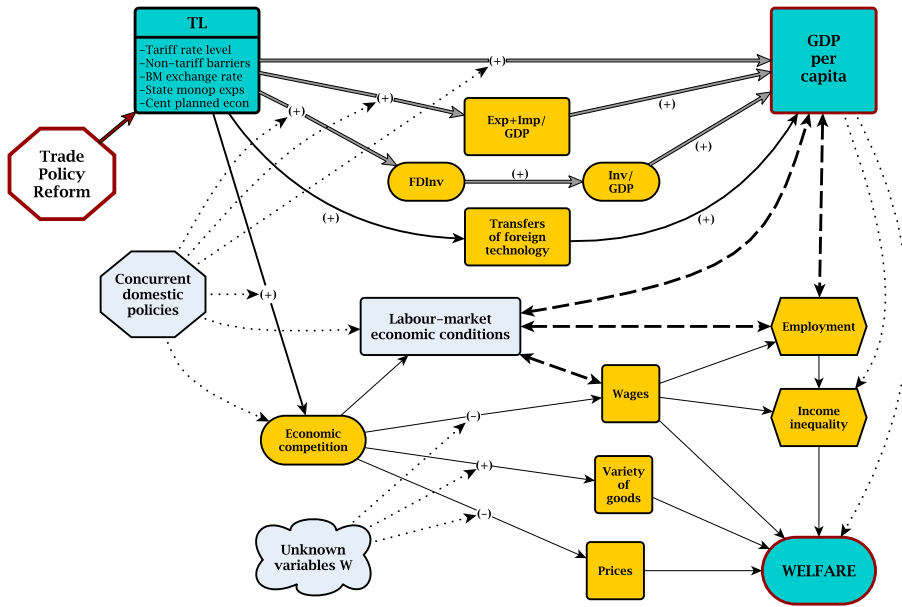


Figure 2. Non-exhaustive set of causal factors and causal connections, based on the tenets of new international trade theory, as an illustration of nodes to be evaluated before a trade policy reform is implemented in a particular country.

predicted implications of the theory will obtain or not in real-life cases will depend (among other things) upon the values of the postulated variables and the efficacy of their posited interconnections in the particular context of application.

As has been argued by several authors (see Driskill 2012), increases in economic growth do not automatically produce improvements in welfare or in welfare distribution. This would still be the case irrespective of any significant effect of TL on GDP per capita, on investment (INV), or on variations in the openness ratio. This is reflected in the causal model of Figure 2. New international trade economics provides some insights on *potential* causal paths that could lead to welfare variations, but the noteworthy point here is that given certain concrete policy aims, all the relevant nodes and causal connections in Figure 2 would have to be evaluated at a local level if one is to have a reliable expectation that the intended welfare-improving policy will be effective. Other theoretical accounts will suggest different potential causal paths that should be checked before trying out any actual policy reform. This is precisely the reason why underlying theoretical accounts should be made explicit in policy deliberations.

5.4. Causal Diagrams as Heuristic Evidential Guidelines for Policy Assessment

Figure 3 displays a non-exhaustive compilation of relevant causal factors and interactions related to trade liberalisation (as derived from the literature reviewed throughout this article).²⁸ All factors included can have a bearing on the effectiveness of a trade

²⁸Different causal diagrams could be designed on the basis of different methodological background assumptions or of different economic theories underlying the empirical research.

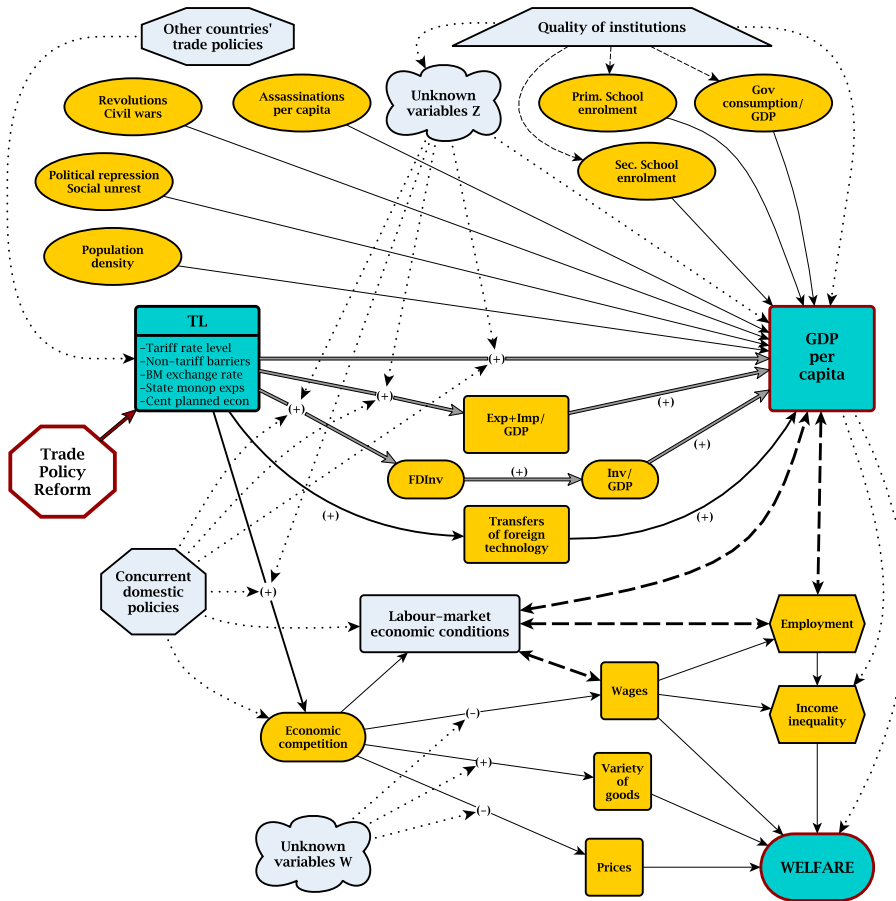


Figure 3. Causal diagram including variables and interconnections that are relevant to the effectiveness of a trade policy reform (in accordance with the background assumptions of the empirical research reviewed in this article).

liberalisation policy reform in a particular country. Careful analysis of the nodes and posited causal connections at the local level would be useful for gathering more reliable evidence to support concrete policy-effectiveness hypotheses.²⁹ All relevant, available and reliable knowledge (interdisciplinary and contextual) about each of the nodes and their interconnections should, if possible, be taken into account before implementing any specific trade policy reform.

The most valuable contribution scientific studies could offer to policy makers is not a stock of well-established lawlike causal generalisations, but rather a clear account of the background assumptions implicit in the research, and the reasons why empirically established causal results might sometimes fail to obtain in concrete and specific real-world situations. In other words, they should provide evidence about the possible disturbing

²⁹Local and regional studies in international economics have become slightly more popular after 2000. These studies show (ex post) the workings of contextual variables that have shown distorting effects in particular regions after trade liberalisation reforms have been implemented (see, e.g., Easterly, Fiess, and Lederman 2003; OECD 2005; Baylis, Garduño-Rivera, and Piras 2012).

and mediating factors, enabling or disabling background characteristics, causal interactions, and any other type of contextual element that could significantly distort the outcomes of real-world socio-economic policy reforms.

6. Conclusions

After analysing some of the most influential studies in mainstream empirical economics about the benefits of international free trade, it can be concluded that there is scientifically valid evidence supporting a broad generalisation about the causal efficacy of trade liberalisation on economic gains. More specifically, there is scientific evidence for positive average causal effects of a variable defined as ‘trade liberalisation’ on other well-known indicators of economic gains (such as, GDP per capita, investment, and so on). Nevertheless, such wide-ranging results are not straightforwardly useful or relevant for informing any specific trade policy decision.

The meaning of the causal hypotheses tested in each particular study depends on the different background assumptions embedded in the econometric techniques employed to test them. Thus, even if the broad causal-efficacy hypothesis that ‘trade liberalisation causes economic gains’ can be taken as scientifically well-established (as many economists do take it), there are at least three methodological aspects of the empirical literature that make the causal results ambiguous or inadequate for policy-making purposes.

First, the variable that is used to measure trade liberalisation is rather vaguely defined from a policy-making point of view. It is usually characterised as a multidimensional variable that allows measurement and econometric estimation. But it is not clear how changes in such a variable would translate into definite real-life policy interventions.

Second, the specific assumptions about the causal criteria and about controlling background conditions, which are implicit in the econometric methods, restrict the inferences that can be drawn about concrete policy applications of the empirical results.

Third, empirical researchers tend to present their scientific results in ways that aim at protecting the general validity of their causal conclusions, e.g. by explaining away observed or potential exceptions, so as to uphold the truth of lawlike generalisations. From a policy-making perspective, a more useful approach would be to tackle straightforwardly the exceptions to the rule, and to offer a clear and explicit characterisation of all the relevant contextual features, background conditions, and disturbing factors that concurrently could enable or constrain the actual occurrence of a causal effect. Such information could provide basic guidelines about which variables are relevant, and how they might affect any intended policy outcomes.

The analysis offered illustrates a mismatch between the kind of evidence typically produced and promoted by economic researchers and the kind of evidence that actually could be useful to policy makers. Most empirical economists working on the research reviewed seem to be aware, to different extents, of their implicit assumptions and the related constraints that such assumptions impose on making policy inferences about specific real-country economies. Nevertheless, the clarifications and qualifications required to properly understand the inferential limits of the empirical results are rarely openly stated or discussed, and thus seldom reach the end-users of scientific research in the policy-making arena.

Acknowledgements

I have received helpful feedback on previous versions of this article at several venues: Philosophy of Social Science Roundtable, Santa Cruz, 2013; European Philosophy of Science Association, Helsinki, 2013; International Network of Economic Method, Cape Town, 2015; European Network for the Philosophy of the Social Sciences, Helsinki, 2016; and the EIPE Research Seminar, Rotterdam, Spring 2017. I am extremely grateful to Alessandra Basso, Marcel Boumans, Aksel Erbahar, Jaakko Kuorikoski, Magdalena Małecka, Caterina Marchionni, Mary S. Morgan, Petri Ylikoski, as well as to two anonymous referees and *ROPE*'s editor Louis-Philippe Rochon, for valuable criticisms and comments that have greatly contributed to improving this article.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

ORCID

Luis Mireles-Flores  <http://orcid.org/0000-0002-4211-1855>

References

- Ades, A. F., and E. L. Glaeser. 1999. 'Evidence on Growth, Increasing Returns and the Extent of the Market.' *Quarterly Journal of Economics* 114 (3): 1025–1045.
- Agosin, M. R., and R. Ffrench-Davis. 1995. 'Trade Liberalization and Growth: Recent Experiences in Latin America.' In *Report on Neoliberal Restructuring, Special Issue of Journal of Interamerican Studies and World Affairs* 37 (3): 9–58.
- Angrist, J. D., and J.-S. Pischke. 2015. *Mastering 'Metrics: The Path from Cause to Effect*. Princeton: Princeton University Press.
- Badinger, H. 2007. 'Has the EU's Single Market Programme Fostered Competition? Testing for a Decrease in Mark-Up Ratios in EU Industries.' *Oxford Bulletin of Economics and Statistics* 69 (4): 497–519.
- Baldwin, R. E., and E. Seghezza. 1996. 'Testing for Trade-Induced Investment-Led Growth.' *NBER Working Paper* 5416. National Bureau of Economic Research, Cambridge, MA.
- Barro, R. J. 1991. 'Economic Growth in a Cross Section of Countries.' *Quarterly Journal of Economics* 106 (2): 407–443.
- Baylis, K., R. Garduño-Rivera, and G. Piras. 2012. 'The Distributional Effects of NAFTA in Mexico: Evidence from a Panel of Municipalities.' *Regional Science and Urban Economics* 42 (1–2): 286–302.
- Ben-David, D. 1993. 'Equalizing Exchange: Trade Liberalization and Income Convergence.' *Quarterly Journal of Economics* 108 (3): 653–679.
- Bernard, A. B., J. Eaton, J. Bradford Jensen, and S. Kortum. 2003. 'Plants and Productivity in International Trade.' *American Economic Review* 93 (4): 1268–1290.
- Best, H., and C. Wolf. 2015. *The SAGE Handbook of Regression Analysis and Causal Inference*. London: SAGE Publications.
- Broda, C., and D. E. Weinstein. 2006. 'Globalization and the Gains from Variety.' *Quarterly Journal of Economics* 121 (2): 541–585.
- Brown, S. 2006. *Myths of Free Trade: Why American Trade Policy Has Failed*. New York: The New Press.
- Brüderl, J., and V. Ludwig. 2015. 'Fixed-Effects Panel Regression.' In *The SAGE Handbook of Regression Analysis and Causal Inference*, edited by H. Best, and C. Wolf. London: SAGE Publications.
- Cartwright, N. D. 2007. *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. Cambridge, UK: Cambridge University Press.

- Chang, H.-J. 2008. *Bad Samaritans: The Myth of Free Trade and the Secret History of Capitalism*. New York: Bloomsbury Press.
- Claveau, F., and L. Mireles-Flores. 2014. 'On the Meaning of Causal Generalisations in Policy-Oriented Economic Research.' *International Studies in the Philosophy of Science* 28 (4): 397–416.
- Coe, D. T., E. Helpman, and A. W. Hoffmaister. 1997. 'North-South R & D Spillovers.' *The Economic Journal* 107 (440): 134–149.
- Daly, H. E. 1996. 'Against Free Trade: Neoclassical and Steady-State Perspectives.' In *The Global Dimension of Economic Evolution*, edited by K. Dopfer. Heidelberg: Physica-Verlag.
- Dean, J. M., S. Desai, and J. Riedel. 1994. 'Trade Policy Reform in Developing Countries since 1985: A Review of the Evidence.' *World Bank Discussion Paper* 267. International Bank of Reconstruction and Development/The World Bank, Washington, DC.
- Dixit, A. K., and V. Norman. 1980. *Theory of International Trade: A Dual, General Equilibrium Approach*. London: Cambridge University Press.
- Dollar, D. 1992. 'Outward-Oriented Developing Economies Really Do Grow More Rapidly: Evidence from 95 LDCs, 1976-1985.' *Economic Development and Cultural Change* 40 (3): 523–544.
- Driskill, R. 2012. 'Deconstructing the Argument for Free Trade: A Case Study of the Role of Economists in Policy Debates.' *Economics and Philosophy* 28 (1): 1–30.
- Dutt, P., D. Mitra, and P. Ranjan. 2009. 'International Trade and Unemployment: Theory and Cross-National Evidence.' *Journal of International Economics* 78 (1): 32–44.
- Easterly, W., N. Fiess, and D. Lederman. 2003. 'NAFTA and Convergence in North America: High Expectations, Big Events, Little Time.' *Economía* 4 (1): 1–53.
- Eaton, J., S. Kortum, and F. Kramarz. 2004. 'Dissecting Trade: Firms, Industries, and Export Destinations.' *American Economic Review* 94 (2): 150–154.
- Edwards, S. 1992. 'Trade Orientation, Distortions and Growth in Developing Countries.' *Journal of Development Economics* 39 (1): 31–57.
- Edwards, S. 1998. 'Openness, Productivity and Growth: What Do We Really Know?' *The Economic Journal* 108 (447): 383–398.
- Falvey, R. 1999. 'Trade Liberalization and Factor Price Convergence.' *Journal of International Economics* 49 (1): 195–210.
- Feenstra, R. C. 1994. 'New Product Varieties and the Measurement of International Prices.' *American Economic Review* 84 (1): 157–177.
- Feenstra, R. C. 2006. 'New Evidence on the Gains from Trade.' *Review of World Economics* 142 (4): 617–641.
- Feenstra, R. C., and H. L. Kee. 2008. 'Export Variety and Country Productivity: Estimating the Monopolistic Competition Model with Endogenous Productivity.' *Journal of International Economics* 74 (2): 500–518.
- Firebaugh, G., C. Warner, and M. Massoglia. 2013. 'Fixed Effects, Random Effects, and Hybrid Models for Causal Analysis.' In *Handbook of Causal Analysis for Social Research*, edited by S. L. Morgan. Dordrecht, NL: Springer.
- Fletcher, I. H. 2010. *Free Trade Doesn't Work: What Should Replace It and Why*. Washington, DC: U.S. Business and Industry Council.
- Frankel, J. A., and D. Romer. 1999. 'Does Trade Cause Growth?' *American Economic Review* 89 (3): 379–399.
- Georgiadis, G., and J. Gräß. 2013. 'Growth, Real Exchange Rates, and Trade Protectionism since the Financial Crisis.' *European Central Bank Working Paper Series* 1618. European Central Bank, Frankfurt am Main, DE.
- Greenaway, D., W. Morgan, and P. Wright. 1997. 'Trade Liberalization and Growth in Developing Countries: Some New Evidence.' *World Development* 25 (11): 1885–1892.
- Greenaway, D., W. Morgan, and P. Wright. 2002. 'Trade Liberalisation and Growth in Developing Countries.' *Journal of Development Economics* 67 (1): 229–244.
- Harrigan, J., and P. Mosley. 1991. 'Evaluating the Impact of World Bank Structural Adjustment Lending: 1980-87.' *Journal of Development Studies* 27 (3): 63–94.

- Harris, R. 1984. 'Applied General Equilibrium Analysis of Small Open Economies with Scale Economies and Imperfect Competition.' *American Economic Review* 74 (5): 1016–1032.
- Harrison, A. 1996. 'Openness and Growth: A Time-Series, Cross-Country Analysis for Developing Countries.' *Journal of Development Economics* 48 (2): 419–447.
- Hausmann, R., D. Rodrik, and A. Velasco. 2008. 'Growth Diagnostics.' In *The Washington Consensus Reconsidered: Towards a New Global Governance*, edited by J. Stiglitz, and N. Serra. New York: Oxford University Press.
- Helpman, E. 1981. 'International Trade in the Presence of Product Differentiation, Economies of Scale, and Monopolistic Competition: A Chamberlin-Heckscher-Ohlin Approach.' *Journal of International Economics* 11 (3): 305–340.
- Hindriks, F. A. 2006. 'Tractability Assumptions and the Musgrave–Mäki Typology.' *Journal of Economic Methodology* 13 (4): 401–423.
- Hitchcock, C. R. 2003. 'Of Humean Bondage.' *British Journal for the Philosophy of Science* 54 (1): 1–25.
- Holland, P. W. 1986. 'Statistics and Causal Inference.' *Journal of the American Statistical Association* 81 (396): 945–960.
- Hoover, K. D., and S. J. Perez. 2004. 'Truth and Robustness in Cross-Country Growth Regressions.' *Oxford Bulletin of Economics and Statistics* 66 (5): 765–798.
- Hummels, D., and P. J. Klenow. 2005. 'The Variety and Quality of a Nation's Exports.' *American Economic Review* 95 (3): 704–723.
- IMF. 1997. *World Economic Outlook*. Washington, DC: IMF.
- Irwin, D. 2009. *Free Trade under Fire*. Princeton, NJ: Princeton University Press.
- James, W. E., S. Naya, and G. M. Meier. 1989. *Asian Development: Economic Success and Policy Lessons*. Madison, WI: University of Wisconsin Press.
- Keller, W. 2004. 'International Technology Diffusion.' *Journal of Economic Literature* 42 (3): 752–782.
- Kincaid, H. 2004. 'Contextualism, Explanation and the Social Sciences.' *Philosophical Explorations* 7 (3): 201–218.
- Kitson, M., and J. Michie. 1995. 'Conflict, Cooperation and Change: The Political Economy of Trade and Trade Policy.' *Review of International Political Economy* 2 (4): 632–657.
- Krueger, A. O. 1983. *Trade and Employment in Developing Countries 3: Synthesis and Conclusions*. Chicago: NBER/University of Chicago Press.
- Krugman, P. R. 1979. 'Increasing Returns, Monopolistic Competition, and International Trade.' *Journal of International Economics* 9 (4): 469–479.
- Krugman, P. R. 1980. 'Scale Economies, Product Differentiation, and the Pattern of Trade.' *American Economic Review* 70 (5): 950–959.
- Krugman, P. R. 1981. 'Intra-Industry Specialization and the Gains from Trade.' *Journal of Political Economy* 89 (5): 959–973.
- Krugman, P. R. 1993. 'The Narrow and Broad Arguments for Free Trade.' *The American Economic Review* 83 (2): 362–366.
- Kuorikoski, J., A. Lehtinen, and C. Marchionni. 2010. 'Economic Modelling as Robustness Analysis.' *The British Journal for the Philosophy of Science* 61 (3): 541–567.
- Lal, D. 2006. *Reviving the Invisible Hand: The Case for Classical Liberalism in the Twenty-first Century*. Princeton, NJ: Princeton University Press.
- Lancaster, K. 1980. 'Intra-Industry Trade under Perfect Monopolistic Competition.' *Journal of International Economics* 10 (2): 151–175.
- Lee, J.-W. 1993. 'International Trade, Distortions, and Long-Run Economic Growth.' *International Monetary Fund Staff Papers* 40 (2): 299–328.
- Levine, R., and D. Renelt. 1992. 'A Sensitivity Analysis of Cross-Country Growth Regressions.' *American Economic Review* 82 (4): 942–963.
- Madsen, J. B. 2007. 'Technology Spillover through Trade and TFP Convergence: 135 Years of Evidence for the OECD Countries.' *Journal of International Economics* 72 (2): 464–480.

- Mäki, U. 1992. 'On the Method of Isolation in Economics.' In *Idealization IV: Intelligibility in Science*, edited by Craig Dilworth, special issue of *Poznan Studies in the Philosophy of the Sciences and the Humanities* 26: 319–354.
- Milberg, W. 1996. 'The Rhetoric of Policy Relevance in International Economics.' *Journal of Economic Methodology* 3 (2): 237–259.
- Mill, J. S. 1874 [1843]. *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*. New York: Harper.
- Mill, J. S. 1874 [1844]. 'Of the Laws of Interchange between Nations; and the Distribution of the Gains of Commerce among the Countries of the Commercial World.' In *Essays on Some Unsettled Questions of Political Economy*. London: Longmans, Green, Reader, and Dyer.
- Milner, C., and P. Wright. 1998. 'Modelling Labour Market Adjustment to Trade Liberalisation in an Industrialising Economy.' *The Economic Journal* 108 (447): 509–528.
- Mireles-Flores, L. 2016. 'Economic Science for Use: Causality and Evidence in Policy Making.' PhD diss., Rotterdam: Erasmus University Rotterdam. <https://repub.eur.nl/pub/93326>.
- Morgan, S. L., and C. Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge, UK: Cambridge University Press.
- Neary, J. P. 2003. 'Monopolistic Competition and International Trade Theory.' In *The Monopolistic Competition Revolution in Retrospect*, edited by S. Brakman, and B. J. Heijdra. Cambridge: Cambridge University Press.
- Ocampo, J. A., C. Rada, and L. Taylor. 2009. *Growth and Policy in Developing Countries: A Structuralist Approach*. New York: Columbia University Press.
- OECD. 1998. *Open Markets Matter: The Benefits of Trade and Investment Liberalisation*. Paris: OECD Publishing.
- OECD. 2005. *Trade and Structural Adjustment: Embracing Globalisation*. Paris: OECD.
- Reinert, E. S. 2007. *How Rich Countries Got Rich ... and Why Poor Countries Stay Poor*. London: Constable & Robinson.
- Reiss, J. 2008. *Error in Economics: Towards a More Evidence-based Methodology*. London: Routledge.
- Reiss, J. 2015. 'A Pragmatist Theory of Evidence.' *Philosophy of Science* 82 (3): 341–362.
- Rodríguez, F., and D. Rodrik. 2001. 'Trade Policy and Economic Growth: A Skeptic's Guide to the Cross-National Evidence.' In *NBER Macroeconomics Annual 2000*, edited by B. Bernanke, and K. Rogoff. Cambridge, MA: MIT Press.
- Rodrik, D. 2006. 'Goodbye Washington Consensus, Hello Washington Confusion? A Review of the World Bank's "Economic Growth in the 1990s: Learning from a Decade of Reform".' *Journal of Economic Literature* 44 (4): 973–987.
- Rodrik, D. 2010. 'Diagnostics before Prescription.' *Journal of Economic Perspectives* 24 (3): 33–44.
- Rodrik, D. 2012. 'Why We Learn Nothing from Regressing Economic Growth on Policies.' *Seoul Journal of Economics* 25 (2): 137–151.
- Rodrik, D. 2015. *Economics Rules: The Rights and Wrongs of the Dismal Science*. London: W. W. Norton and Co.
- Romer, P. M. 1990. 'Endogenous Technological Change.' *Journal of Political Economy* 98 (5): S71–S102.
- Sachs, J. D., and A. Warner. 1995. 'Economic Reform and the Process of Global Integration.' *Brookings Papers on Economic Activity* 1995 (1): 1–118.
- Smith, A., and A. J. Venables. 1988. 'Completing the Internal Market in the European Community: Some Industry Simulations.' *European Economic Review* 32 (7): 1501–1525.
- Spanos, A. 2015. 'Revisiting Haavelmo's Structural Econometrics: Bridging the Gap between Theory and Data.' *Journal of Economic Methodology* 22 (2): 171–196.
- Steedman, I. 1979. *Fundamental Issues in Trade Theory*. London: Macmillan Press.
- Summers, R., and A. Heston. 1991. 'The Penn World Table (Mark 5): An Expanded Set of International Comparisons 1950–88.' *Quarterly Journal of Economics* 106 (2): 327–368.

- Taylor, L. 2010. *Maynard's Revenge: The Collapse of Free Market Macroeconomics*. Cambridge, MA: Harvard University Press.
- Trefler, D. 2004. 'The Long and Short of the Canada-U.S. Free Trade Agreement.' *American Economic Review* 94 (4): 870–895.
- Tsunekawa, K. 2019. 'Emerging States in Latin America: How and Why They Differ from Their Asian Counterparts.' In *Emerging States at Crossroads*, edited by K. Tsunekawa, and Y. Todo. Singapore, SG: Springer.
- Wacziarg, R. 2001. 'Measuring the Dynamic Gains from Trade.' *The World Bank Economic Review* 15 (3): 393–429.
- Wacziarg, R., and K. Horn Welch. 2008. 'Trade Liberalization and Growth: New Evidence.' *The World Bank Economic Review* 22 (2): 187–231.
- Williamson, J. 2000. 'What Should the World Bank Think about the Washington Consensus?' *World Bank Research Observer* 15 (2): 251–264.
- Winters, L. A., N. McCulloch, and A. McKay. 2004. 'Trade Liberalization and Poverty: The Evidence So Far.' *Journal of Economic Literature* 42 (1): 72–115.
- Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.
- World Bank. 1993. *Report on Adjustment Lending III*. Washington, DC: The World Bank.
- Young, S. 1996. 'Political Economy of Trade Liberalization in East Asia.' In *The World Trading System: Challenges Ahead*, edited by J. J. Schott. Washington, DC: Institute for International Economics.