

*Mariusz Maziarz**

IT'S ALL IN THE EYE OF BEHOLDER¹

The article aims at delivering a fresh look at the dependency of econometric results on research methods. The author discusses the Reinhart-Rogoff controversy and the literature on expansionary fiscal contractions with a view to argue that econometricians can obtain recalcitrant (inconsistent) results despite using similar datasets. Case studies are used to show the dependence of causal claims on econometric techniques. The choice of a weighting scheme or a method identifying the timing of historical contractions determines the obtained results. Considering the strong policy-orientation of the case studies, the article delivers evidence that economic policy-making is susceptible to the research methods used for the production of evidence. This study supports the conventionalist and constructivist stance in the philosophy of econometrics.

Keywords: debt-to-GDP, expansionary fiscal contractions, economic methodology, constructivism, empirical macroeconomics, economic policy-making

JEL Classifications: B41, N10, E62

DOI: 10.15611/aoe.2019.2.13

1. INTRODUCTION

Although the recent economic slump and the subsequent global recession caused a passing convergence of both academic economists' and public policy practitioners' heterogeneous points of view into the interventionist's one, the opinions among economists are usually evenly divided. For instance, Keynesians and neoclassical economists have argued in favor of either constant or immediately adjustable prices for close to 80 years. Much water has passed under the bridge since the publishing of *The General Theory of Employment, Interest and Money* but the two schools of economic thought still cannot agree on a number of substantial facts.

Growing government debts and economic recoveries have prompted economists address the question if (or, possibly, how deeply is more accurate) governments should be troubled by the enormity of the deficits.

* The Faculty of Economic Sciences, Wrocław University of Economics and Business.

¹ The research was supported by National Science Centre, Poland under grant no. 2015/19/N/HS1/01066. The author acknowledges Ph.D. scholarship from the National Science Centre, Poland, under grant no. 2018/28/T/HS1/00007.

Debt overhangs caused by multi-billion bailouts and reduced tax incomes have motivated the recent intensification of research on the relation between government debt and the pace of economic growth and the possible mechanisms connecting the two (e.g. Kersan-Skabic 2016). One of the soundest cases of this wave of studies is Reinhart and Rogoff's (2010a) cliometric study delivering evidence for the negative influence of public debt on economic development. Policymakers used these results as evidence for spending cuts.

Another tough question that demands an immediate answer is defining when governments should cut expenditure or, to put it in other words, what are the outcomes of austerity at the Treasury. Is reducing government debt always supporting economic growth (by lowering interest rates, for instance) or does it lessen the output (due to, e.g. reducing aggregated demand) in the short term? Understanding this critical issue is helpful in deciding whether governments should, as the Irish proverb says, *fix the roof when the sun is shining* by only introducing austerity programs after the recession ends or else a delay is unnecessary because, on the one hand, it makes the debts expand over time and, on the other, reducing public deficit is expansionary even if undertaken during a recession. As long as the theoretical mechanisms are still unknown and experimentation in the macroeconomic realm is impossible due to epistemic reasons, the crucial decisions depend on analyzing observational data. However, the policymaking advice delivered by different teams of econometricians differs vastly.

The article's primary objective is describing the process of constructing econometric 'observations' by choosing the research methods. Considering that econometric results are considered on the ground of economic methodology as tests for theory (e.g. Friedman 1953, Blaug 1992), they seem to be taken as more confident than theorizing and are directly applied as evidence for policymaking (Ryan 2011). However, a closer study of the dependence of the results on research methods sheds light on the lack of certitude of econometric results and its production by research methods.

Furthermore, the author wants to understand how econometric observations develop. The case study based on the threshold hypothesis focuses on the broadly known case of the two inconsistent findings obtained by Reinhart and Rogoff (2010a) and Herndon, Ash and Pollin (2014)². To the best of our knowledge, the voices committed to the methodology applied in both the journal literature (for instance Clemens 2015; Muslu et al. 2015;

² RR and HAP henceforth, accordingly.

Dafoe 2014; Grimson 2014; Reiss 2014) as well as the popular press (e.g. Stevenson and Wolfers 2013) focus on the necessity of replicating economic analyses, making databases publicly available and underline the spreadsheet error committed by RR. Some commentators (e.g. Okalow 2013; Wray 2013) explicitly accused RR of fudging and flubbing empirical work for the sake of getting the desired results what, as is shown below, completely misses the point. Section 2 analyzes their methodologies and gives a brief review of the diverse literature on the threshold hypothesis in order to justify the point of view that the difference between the RR and HAP results is mostly driven by various methodological choices and the alternatives are justified to a similar degree (cf. Maziarz 2017; Bitar et al. 2018). The case described in Section 3 is based on two contradictory articles on the expansionary fiscal contraction hypothesis. Guajardo, Leigh, and Pescatori (2010)³, in contrast to Alesina and Ardagna (2009)⁴, refute the positive short-term effects of austerity at the treasury.

The two pairs of economic results are contrary in the sense that they oppose each other: they either differ on the existence of a connection between the two variables (RR vs. HAP) or disagree on the sign of relation (GLP vs. AA). These case studies also show that the differentiation of findings is caused by different methodological choices (a method of identifying the years of fiscal contractions and the choice averaging scheme) that, as the author argues, are arbitrary. In other words, none of the methods is superior on the ground of economic methodology. In Section 4, the author generalizes the conclusions from the discussed case studies and considers how econometric observations are developed in order to deliver an explanation why there are groups of economists who hold opposite points of view. Finally, the main conclusions are given and indicate the areas of further research.

2. THE THRESHOLD HYPOTHESIS

The most often quoted articles are usually the least read ones. Therefore a review of the RR and HAP methodologies that led to the contradictory findings is necessary. In addition to familiarizing the reader with research methods, the author also discusses arguments in favor of the alternative methodologies. RR based their research on their own database on gross

³ GLP, henceforth.

⁴ AA, henceforth.

central government debt spanning approximately two hundred years. The full dataset consists of over 3700 country/year observations that were classified into one of the four groups (buckets) where the debt/GDP ratio value equaled less than 30%, 30 to 60%, 60 to 90% and more than 90%. RR considered three samples: the long one, post-war emerging markets and post-war developed countries.

HAP aimed at replicating this study. However, their research does not fulfill the usual understanding of replication. Due to the different methodological choices, it should be called a robustness test instead (Clemens 2015). This article considers the problem of choosing an averaging scheme in detail. Most specific cross-country time series contain debt/GDP ratios that are classified into any one of the above groups more than once. For instance, the Swiss total gross central government debt/GDP time series published on Carmen Reinhart's webpage⁵ consists of 130 instances. Four of the yearly observations fell into the 60-90 percent limits. On the other hand, there is only one value of the debt/GDP ratio classified in this bucket in the case of Norway. In order to clearly explain the methodological choice of an averaging scheme, the author assumes that the database consists of these two countries only and is interested in the question of what is the average economic growth of countries during the years when their debt/GDP ratio fell within the 60-90 percent range. In this case, there are only five observations to be averaged, but the number of possibilities of how the average GDP growth can be calculated is infinite. Let us consider the following two examples:

$$X_1 = \frac{GDP_{No1946} + GDP_{Ch1944} + GDP_{Ch1945} + GDP_{Ch1946} + GDP_{Ch1947}}{5}, \text{ (HAP)}$$

$$X_2 = \frac{GDP_{Ch1944} + GDP_{Ch1945} + GDP_{Ch1946} + GDP_{Ch1947} + GDP_{No1946}}{4} \cdot \frac{1}{2}. \text{ (RR)}$$

The first method (applied by HAP) assigns the same influence to every country/year observation. Hence, every one of the four values of the Swiss growth observed in the considered bucket influences the average in the same way, even though the reason of the debt overhang is easy to guess, taking into account that these are the years of the Second World War and the following years. Bearing in mind that the events that caused the debt

⁵ Carmen M. Reinhart, <http://www.carmenreinhart.com/data/browse-by-country/>, access: 17th September 2015.

overhang are unlikely to be repeated (and they certainly do not within the period considered), should every one of the four values of the Swiss GDP growth influence the average to the same degree as the Norwegian one?

In order to reduce the influence of countries characterized by the consecutive high levels of the debt/GDP ratio on the value of the X_1 , an alternative averaging scheme X_2 can be calculated in two steps. First, the GDP growth rate of every country in a bucket is arithmetically averaged. Second, the above calculated average growth rates of every country are used to quantify arithmetic mean. In this case, every country influences the bucket's average value of growth in the same way, no matter how many country/year observations belong to the bucket under consideration. In addition to the two extreme cases, there is an infinite set of averaging schemes based on different weighting methods.

RR decided to use the second averaging method (X_2). On the other hand, HAP favored the unweighted method (X_1) and called the contrary choice a *non-standard* one even though they coined a reason for RR's decision: *possible within-country serially correlated relationships could support an argument that not every additional country-year contributes proportionally additional information* (HAP, pp. 7-8). The debt/GDP ratio is known to be serially correlated due to the so-called debt overhangs, where lower economic growth causes higher public indebtedness (Dafermos 2015; Reinhart, Reinhart and Rogoff 2012).

Kumar and Woo (2010) delivered empirical evidence in favor of the hypothesis that initial high public debt determines the subsequent growth. Assuming that the periods of high public indebtedness are caused by negative shocks which is grounded in an observation of Reinhart, Reinhart and Rogoff (2012) who pointed out that debt overhangs build up mostly during recessions but last much longer, country/year observations following one another should not influence the average to the same degree. Pescatori, Sandri and Simon (2014) delivered evidence in favor of the RR's averaging scheme finding that rising debt/GDP ratio is correlated to lower economic development: *GDP growth averages around 2 percent in countries with debt below 90 percent, and tumbles to about -2 percent in countries whose debt ratio increases above that level* (p. 7). Therefore, the unweighted average (X_1) might underestimate the influence of high indebtedness on growth.

On the other hand, RR's averaging method is also criticized. The main argument against the weighted average (X_2) states that it makes a single country/year observation influence the overall findings in the same way as a multi-year average of the GDP growth of a country that fell within

a considered bucket. Hence RR's averaging scheme seems (e.g. for HAP) intuitively invalid because it leads to situations when a single country/year observation strongly influences the average GDP growth of a whole bucket (as in the case of New Zealand). The high serial correlation of debt leads to situations when one adverse event (a war or a significant economic recession) that raises the levels of debt-to-GDP ratio strongly influences the estimates and biases the results. The opponents of RR highlighted that even though serial correlation in the debt/GDP ratio exists, it does not justify equaling the experience of – for example – Greece (19 years in the highest debt/GDP ratio bucket; 2.9% average GDP growth) and New Zealand (one year; 7.6% GDP slump).

Moreover, both these averaging schemes in common are fallacious due to the explicit assumption that the average pace of economic growth is constant among different countries which is called in the econometric literature *country fixed effects*. Taking into account the differences among countries, considering some different (country-specific) threshold levels seems to be more justified (Bell, Johnston and Jones 2014; Egert 2012; Kourtellos, Stengos, Ming-Tan 2012).

In light of the above arguments, whose weights seem to be similar⁶, both methods can be said to be equally justified (or equally flawed). In fact, Hamilton (2013), an econometrician focused on time series analysis, supported this point of view on the arbitrariness of choosing the averaging method and claimed that suggesting *that there is some deep flaw in the method used by RR or an obvious advantage to the alternative favored by HAP is in my opinion quite unjustified*. Moreover, Reinhart and Rogoff (2010b), before the affair started and the criticism appeared in the literature, published an essay on the threshold hypothesis where they explicitly stated that the methodological choices made in the field of cliometrics are not grounded in econometric knowledge: *Those who have done data work know that mapping vague concepts like “high debt” or “overvalued exchange rates” into workable definitions requires arbitrary judgments about where to*

⁶ It is certainly impossible to measure weight or strength of evidence in favor of the two averaging schemes since being convinced is subjective (or intersubjective, at best). In addition, the author argues below that the points of view on methodological commitments are shaped by the presuppositions that economists bring to their work rather than the evidence. Therefore, further discussion is based on the assumption grounded in the literature reviewed above indicating that the weight of arguments is similar and the methods are equally justified. However, the point of view presented hitherto by the commentators that only one of the averaging schemes is right and the other is flawed is surely not legitimized.

draw lines; there is no other way to interpret the facts and inform the discussion.

Even though choosing one of the two averaging schemes may be called arbitrary (i.e., none of the methods is more grounded in econometric knowledge), the decision influences the values of the means. The differences between the averages calculated with RR's or HAP's methodology in every bucket spans from circa 0.1 percentage point up to 0.3 percentage point in the group that got most attention, i.e. where the debt/GDP ratio exceeds 90 percent, see Table 1. Moreover, the influence of the averaging methodology is multiplied by other methodological choices that are described below. When the two average schemes are applied to the data including the spreadsheet error and the New Zealand data exclusion, the difference equals 1.7 percentage point for the over 90 percent bucket (HAP, p. 21).

Table 1

The differences between RR and HAP averaging methods when applied to the same database

Debt/GDP ratio	Below 30 percent	30 to 60 percent	60 to 90 percent	Over 90 percent
RR'smethod	3.98	3.00	3.05	1.96
HAP'smethod	4.18	3.08	3.30	2.21

Source: own calculations based on data from Table A-1 (HAP).

There are two additional errors pointed out by HAP. In fact, one of them, the spreadsheet error, can legitimately be called an error. The second one is data omission. RR committed the spreadsheet error that excluded from the summary statistics of five countries (Australia, Austria, Belgium, Canada, and Denmark). The coding error influenced averages in every sample and across all the debt/GDP ratio buckets (Reinhart and Rogoff 2013b). The difference in means caused by the exclusion equals 0.3 percentage point in the highest debt/GDP ratio bucket and fell within the 0.1 to 0.2 percentage point range in the remaining groups.

Facing the accusation of excluding the New Zealand (and Spain which HAP did not mention) data on purpose in order to get the desired result, Reinhart and Rogoff(2013a) reply that there are two GDP estimates accessible in the case of New Zealand (delivered by Angus Maddison's Database and the New Zealand Historical Statistics). Furthermore, at the time when the research was conducted, there were no trustworthy estimates of Spanish GDP before 1960. In this case, similar to the averaging method, there was no mistake and the correct choice was made, since the knowledge

was limited and some arguments supported excluding these countries and others including the uncertain estimates. In fact, RR had the possibilities: either excluding New Zealand from the database or basing the research on one of the estimates. At the time of conducting the research, there was no ground to base the choice on. Considering the influence⁷ of the decision on the results, the exclusion seems to be justified. However, there are arguments in favor of doing the opposite.

Another methodological choice lacking attention that may influence the overall result of research on the threshold hypothesis is deciding what kind of debt should be used in the calculations. RR chose gross central government debt most likely due to the data accessibility. The datasets of total gross general government debt, or public plus private gross external debt include a shorter time series. However, similar research based on public debt defined in another way (e.g. as a total gross external debt) might lead to obtaining slightly different results. Egert (2015, p. 3765) calculated average GDP growth from 1960 to 2009 for the four buckets of the debt/GDP ratio defining it as either general government debt or central government debt. The differences between the means amounted up to 0.5 percentage point.

Summing up, the spreadsheet error hardly influenced the GDP growth estimate (up to 0.3 percentage point). Instead, the difference between the RR and HAP findings resulted from the methodological choices described above (cf. Maziarz 2017):

1. averaging schemes (up to 0.3 percentage point when HAP's data and their other methodological commitments are quantified but multiplied up to 1.7 percentage points when considering RR's methodological decisions);
2. whether the New Zealand GDP estimate should be included (up to 0.6 percentage point);
3. what kind of debt the research should be based on (up to 0.5 percentage point).

Therefore, it is justified to conclude that the contrariness of the RR and HAP results is grounded in the different methodological decisions (i.e. not mistakes). It is undoubtedly impossible to measure the weight of the above-discussed arguments in favor of RR or HAP methodology. However, one plainly can not state that a set of choices is right and the contrary one is

⁷ The New Zealand exclusion reduces the average GDP growth estimate in the highest debt/GDP bucket of 0.3 percentage point (HAP). On the other hand, including one of the two estimates causes the estimate to differ from 1.9% (the Maddison database) to 2.5 percent (the New Zealand Historical Statistics; Reinhart and Rogoff 2013b).

wrong. One cannot state that, for instance, the unweighted averaging scheme is justified and call the weighted one a mistake. Hence, it is essential to ask the question what makes one group of economists support the threshold hypothesis and the other oppose it? What are the causes of the differentiation? The author attempts to answer these questions in Section 4.

Table 2
Selected econometric research on the threshold hypothesis

Authors (year of publication)	Findings
Egert (2015)	1. The lower threshold at 20%, higher somewhere between 55% and 130%. 2. One threshold well below 60%.
Baum, Checherita-Westphal, Rother (2012)	1. Positive relation up to 67% of the ratio, high levels of debt/GDP ratio (over 95%) hamper economic growth.
Kourtellos, Stengos, Ming Tan (2012)	1. Countries with weak democratic regimes suffer from high indebtedness more.
Afonso and Jalles (2011)	1. The significant 90% threshold. 2. Increases in the ratio influence more strongly countries with higher debts.
Goulas and Zervoyianni (2013)	1. High debt levels hamper GDP per capita growth only in cases of high uncertainty.
Minea and Parent (2012)	1. The threshold around the debt/GDP ratio of 115%.
Pescatori, Sandri, Simon (2014)	1. Debt trajectory influences subsequent economic development to a higher degree than the debt/GDP ratio.
Lee et al. (2014)	1. The threshold is at the level of 30% of the debt/GDP ratio.
Egert (2012)	1. The threshold levels are not robust. 2. Assuming a linear relation, a ten percentage point increase in the public debt ratio is associated with 0.1 to 0.2 percentage point lower economic growth.
Berhardt and Presbitero (2013)	1. The threshold levels depend on the methods of estimation and model specification. 2. There possibly are thresholds at 52%, 75% and 90% of the debt-to-GDP ratio.
Kumar and Woo (2010)	A possible threshold at the level of 90%. Assuming a linear relation, a ten percentage point increase in the initial debt-to-GDP ratio is associated with a slowdown in annual real per capita GDP growth of around 0.2 percentage point per year, with the impact being somewhat smaller in advanced economies (0.15).

Source: Maziarz (2016).

The author's point of view on the similar justification of both methodologies can be supported by the fact that the findings of econometric analyses aimed at refuting the threshold hypothesis are also divided (cf. Table 2). Egert (2012) concluded his econometric investigation on the threshold hypothesis by writing that the parameters' estimations are unstable, strongly depend on the choice of countries, threshold levels, and model parameterization (p. 15). Do methodological choices influence the results one obtains in the research on the threshold hypothesis only? In order to refute such doubts, the author presents another case study of two contradictory econometric results in Section 3.

3. THE EXPANSIONARY FISCAL CONTRACTION HYPOTHESIS

Assuming that the views of both economists and economic policy practitioners are in favor of the threshold hypothesis (which was the case in the debate a few years ago, namely from 2010 to 2013) and considering the then-current debt/GDP ratios at the historically highest levels, the decision was to be made when the debts should be reduced. The solution seems to be given by the research on the expansionary fiscal contraction hypothesis. Similarly to the previous case study, the views on whether the fiscal contraction is expansionary are strong on both extremes. Contrary to the research on the threshold hypothesis, the difference is caused by a single methodological choice⁸.

Despite the lack of coherence in the empirical findings, they constitute the only evidence in favor (or against) the expansionary fiscal contraction hypothesis available as long as the economic theory is split. Keynesians, believing in constant prices and wages in the short term, say that fiscal contraction causes a temporary economic slump due to a reduction in aggregate demand. Neoclassical economists, on the contrary, argue that fiscal contractions might be expansionary because of raising customers' wealth (due to a reduction of taxation and risk premium) and reducing wages, which positively influences national competitiveness (Alesina and Perotti 1996).

Research on the expansionary fiscal contraction hypothesis is aimed at measuring the influence of government spending cuts on economic development. The first step of the analysis is indicating the years of fiscal contractions in order to estimate their average influence on output. There are

⁸ Further on, the author focuses on the methodological choice leading to the opposite results. For a general methodological discussion, see e.g. Perotti (2011).

roughly two ways of finding when governments have cut spending. On the one hand, GLP, in order to identify the years of austerity in the treasury, applied the methodology developed by Romer and Romer (2007) who based their research on the effects of fiscal shocks on an analysis of narrative records. In order to implement the action-based approach, GLP analyzed *OECD Economic Surveys, IMF Staff Reports, IMF Recent Economic Developments reports, country budget documents and additional country-specific sources* (p. 97).

On the other hand, AA took into account changes in cyclically adjusted primary balance (CAPB)⁹ in order to define fiscal impulses. In particular, there are some approaches to cyclical adjustments. The simplest one is omitting this step and considering changes in the primary balance (Alesina, Perotti 1995). However, this approach does not warrant that changes in primary balance reflect fiscal stimuli and contractions since the primary balance variance can equally often be driven by the business cycle.

Therefore, AA decided to cyclically adjust the primary balance with methodology elaborated by Blanchard (1990) where CAPB is calculated as if the unemployment rate observed in a previous year were unchanged. There are more sophisticated methods developed by the IMF and the OECD. AA consider them as inappropriate, being based on too many strict assumptions about fiscal multipliers. The IMF methodology is based on the concept of the base year, see Table 3. It is essential to highlight that different

Table 3

The approaches to fiscal impulses identification

Primary balance changes	$FI = (g_t - t_t) - (g_{t-1} - t_{t-1})$
Blanchard (1990)	$FI = [(g_t(U_{t-1}) - t_t) - (g_{t-1} - t_{t-1})]$
OECD	$FI = [(G_t - T_t) - (G_{t-1}(1 + \hat{y}_t) - T_{t-1}(1 + y_t))] / Y_{t-1}$
IMF	$FI = [(G_t - T_t) - (G_0(1 + \hat{y}_t) - T_0(1 + y_t))] / Y_{t-1}$

Note: G_t – total current expenditure plus gross capital accumulation less interest payments; T_t – total revenues; g_t ; t_t – expenditure and revenues (accordingly) as a share of GDP; Y_t – nominal GDP; y_t – the rate of growth of nominal GDP; \hat{y}_t – the rate of growth of nominal potential GDP; G_0 – value of G in a base year when actual output is equal to potential output; T_0 – revenues in a base year; U_t – unemployment rate.

Source: Alesina and Perotti (1995).

⁹ Government net borrowing, excluding interest payments on consolidated government liabilities.

approaches lead to nontrivial differentiation of findings¹⁰ (Alesina and Perotti 1996). Some economists argued in favor of less orthodox approaches, e.g. considering military spending, which is said to reflect fiscal stimuli due to being unrelated to cyclical changes in output. AA took into account fiscal impulses larger than 1.5 percent of GDP.

Table 4

A comparison of the years after 1980 when fiscal contractions were identified with the two approaches

Country	Action-based methodology (GLP)	CAPB-based approach (AA)
Australia	1986; 1987	1987; 1988
Austria		1984; 1996; 1997; 2005
Belgium	1982; 1983; 1987; 1993	1982; 1984; 1987; 2006
Canada		1981; 1986; 1987; 1995; 1996; 1997
Denmark	1983; 1984; 1985; 1986	1983; 1984; 1985; 1986; 2005
Finland	1992; 1993; 1994; 1996; 1997; 1998	1981; 1984; 1988; 1994; 1996; 1998; 2000
France		1996
Germany	1997	1996; 2000
Greece		1986; 1991; 1994; 1996; 2005; 2006
Ireland	1982; 1983; 1987; 1988; 2009	1984; 1987; 1988; 1989; 2000
Italy	1992; 1993; 1995; 1997	1980; 1982; 1990; 1991; 1992; 1997; 2007
Japan	1997	1984; 1999; 2001; 2006
Netherlands		1983; 1988; 1991; 1993; 1996
New Zealand		1987; 1989; 1993; 1994; 2000
Norway		1980; 1983; 1989; 1996; 2000; 2004; 2005
Portugal	1983; 2002	1982; 1983; 1986; 1988; 1992; 1995; 2002; 2006
Spain		1986; 1987; 1994; 1996
Sweden	1983; 1993; 1995; 1996; 1997	1981; 1983; 1984; 1986; 1987; 1994; 1996; 1997; 2004
United Kingdom	1981; 1997	1982; 1988; 1996; 1997; 1998; 2000
United States	1991	

Source: GLP, Table 3.4; AA, Table A1.

¹⁰ However, the differences are lower than in case of the CAPB and narrative-based methodologies comparison.

Despite the differentiated results, both methods are supported by some evidence. The main arguments are drawn by opponents of the expansionary fiscal contraction hypothesis, who point out that changes in CAPB do not reflect discretionary policy changes but an economic cycle or other events. For instance, a rise in asset prices that makes government's revenue higher might be interpreted as a fiscal contraction since the primary balance increases. Gaujardo, Leigh, and Pescatori (2011), one year after *The Economic Outlook* containing GLP had been published, analyzed both methodologies and argued that the CAPB-based methodology is *likely to bias the analysis towards downplaying the contractionary effects of deliberate fiscal consolidation* (p. 3). GLP conclude the review of both methodologies giving an example of omitting years when fiscal consolidations are followed by a negative shock. Hence, the supposed bias is caused by excluding the contractions followed by an economic slump.

The opposing methodology, based on the narrative approach, is found to be fallacious by some economists, too. The criticism is less frequent because the method is discussed and employed not so often as the one previously discussed. Perotti (2011) argues that *discretionary fiscal consolidations are often smaller than estimated in the past* (p. 4). The argument against the action-based methodology is the fact that governments are likely to change their spending plans in reaction to current economic events. For instance, if a government abandons its austerity plans in order to reduce the influence of an unpredicted negative shock on the economy, then the CAPB-based approach may be more reliable and the narrative methodology biased toward overestimating contractionary effects¹¹.

Similarly to the previous case study, both methods of indicating when austerity at the Treasury takes place are, taking into account the argumentation discussed above, justified to a similar degree. Comparing the weight of arguments certainly extends the scope of this article. However, the current research practice supports the author's point of view. Namely, both methods are still widely used in the research on the expansionary fiscal contraction hypothesis. For instance, Afonso and Martins (2014) applied the CAPB-based methodology to deliver evidence in favor of the hypothesis based on the example of 14 European Union countries during 1970-2012. Afonso and Jalles (2014) took advantage of using both methodologies and

¹¹ In this case, the CAPB-based methodology would not indicate the considered year as one when government conducts spending cuts. On the other hand, the narrative approach would do, even though austerity plans were cancelled during the considered year.

showed that the obtained results differ. Some authors, in spite of pointing out the shortcomings, apply the CAPB-based methodology to identify the years of fiscal contractions. Dallepiane-Avellaneda (2014) conducted the most up-to-date review of the research on expansionary fiscal contraction hypothesis. In fact, even scholars connected to IMF, the institution employing Guajardo, Leigh and Pescatori, stated that *empirical findings tend to differ, depending on the method used to identify consolidation episodes and the “traditional” approach based on changes in the CAPB is still the most commonly used* (Baldacci et al. 2013, p. 5). To sum up, the case study analysis showed that the outcome of research on the expansionary fiscal contraction hypothesis is determined by the methodological choice, where both alternatives can be said to be justified to a corresponding degree.

4. CONSTRUCTING A CAUSAL CLAIM

The above-discussed case studies show that there are groups of economists who consider different methods as appealing or right and arrive at contrary results. Moreover, the comments on the Reinhart-Rogoff controversy indicate that economists engaged in the discussion either support a set of methodological choices that lead to the finding supporting the threshold hypothesis or endorse the alternative one and hence find no threshold. Similarly, the alternative methods of indicating when austerity at the treasury takes place lead to obtaining the opposite conclusions, but the cliometricians committed to the research on the expansionary fiscal contraction hypothesis seem to support one of the methodologies and disregard the other. However, the exact analysis of the arguments in favor of the methodological choices shows that the weight of arguments (to use the Keynesian saying) is similar. In this section, the author rationally reconstructs the explanations already delivered by the methodologists interested in the Reinhart-Rogoff controversy, shows that they lack descriptive adequacy and offers an argument in support of the constructivist/philosophy of econometrics.

Considering the similarly justified evidence in favor of the opposing causal claims regarding the debt-to-GDP threshold and expansionary fiscal contraction hypotheses, there are two competing stylized facts in the literature that describe either the existence or non-existence of the threshold and the influence of fiscal contraction on economic growth. What made one group of economists support a viewpoint on the required policy and another

oppose it? As mentioned in the introductory section, there are two points of view on the Reinhart-Rogoff controversy present in the literature. On the one hand, the commentators underline the spreadsheet error and other “flaws” committed by RR and advise replicating economic research (Clemens 2015; Muslu et al. 2015; Dafoe 2014; Grimson 2014; Reiss 2014; Stevenson and Wolfers 2013). On the other hand, there are voices of criticism from the perspective of the “economics of economics”. In this approach, RR are accused of committing the spreadsheet error and methodological decisions in order to get the desired result (Okalow 2013; Wray 2013).

The first way of explaining the contradictory points of view on the threshold hypothesis (which we will call the standard approach since it is the most popular in the literature) is based on the unspoken assumption that the cliometric methodology is constant and widely accepted by the economists devoted to this empirical discipline. In this framework, the spreadsheet error (the most sound) and the other methodological choices pointed out by HAP as inappropriate, i.e. the exclusion of New Zealand and the unusual averaging scheme, are indicated as the leading cause of the differentiated results, see Figure 1.

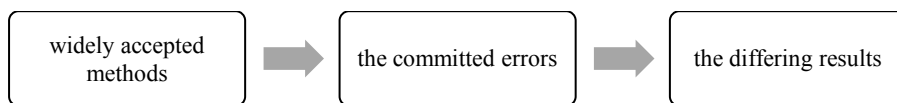


Fig. 1. The standard view on econometrics

Source: own work.

The second explanation (grounded in the economics-of-economics perspective) present in the commentary literature is based on accusing RR of producing the causal claim aimed at convincing other economists or economic policy-makers to the stance supporting the limited role of government in economics. In fact, their article turned out to be very popular and influential since it was cited not only by academics, but also by some government reports worldwide (Maziarz 2017). In accordance with the economics of economics perspective, RR faced institutional constraints and incentives during their scientific work which forced them to produce the desired result. In this case, the methodological choices are supposed to be undertaken with a view to producing evidence in favor of the 90% threshold hypothesis. In line with this perspective, the causal chain usually leading from methods through employing them to the research to the conclusions is

reversed, i.e. the economists are suspected to define a conclusion due to non-scientific reasons and choose the methods that make reaching it possible, see Figure 2.

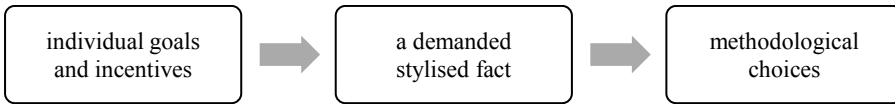


Fig. 2. The economics-of-economics perspective.

Source: own work.

However, the above-discussed case study based on RR and HAP articles showed that the spreadsheet error influenced the results only to a limited degree. On the contrary, the inconsistent results occurred due to the alternative methodological choices that, as argued above, can accurately be said to be similarly justified. Hence, the two explanations present in the literature lack descriptive adequacy since the commentators have not considered the fact that the various methodologies are justified to a corresponding degree. On the contrary, they literally admitted that only one of the diverging methods is right.

However, the above case studies showed that the considered methodological decisions (such as choosing an averaging scheme and a method of indicating when governments cut their spending) can legitimately be called conventions that, as Fleck (1979, p. 9) put it, in spite of seeming equally possible from the point of view of logic, are felt to vary in utility for scientists. According to Fleck's constructivism, the contradictory findings arrived at by RR and HAP can be explained by the fact that the authors of these articles belong to different thought styles¹². In line with the constructivist approach, the results of econometric modeling develop as follows. First, some factors such as educational background, exchanging ideas with co-workers and other cliometricians, etc. made RR and HAP (also

¹² Ludwig Fleck, a microbiologist and a philosopher of science who published his thoughts on the philosophy and sociology of science in 1935 in Germany, investigated the development of medical theories on syphilis and causes of differentiation of anatomic observation. He coined the term "thought style" and defined it as a way of thinking that determines the formulation of every concept (Fleck 1979, p. XXVII). It is characterized by common features in the problems of interest to a thought collective, by the judgment which the thought collective considers evident, and by the methods which it applies as means of cognition. The thought style may also be accompanied by a technical and literary style characteristic of the given system of knowledge (Fleck 1979, p. 99).

AA and GLP) develop divergent presuppositions on methodology. Second, even though the different methods are likely to be equally justified, they do not seem so for the economists belonging to two different thought styles, i.e. RR with HAP, and GLP with AA perceive the methods employed by them as justified and the alternative ones as erroneous. Third, as shown above, applying the alternative methods leads to arriving at contrary findings, see Figure 3.

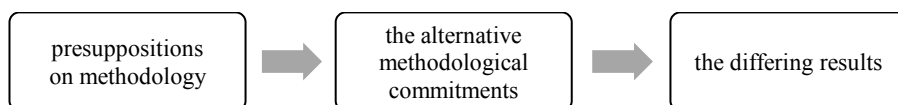


Fig. 3. A constructivist/conventionalist approach

Source: own work.

The constructivist stance on econometric observation has two main advantages in comparison to the views discussed previously. First, the author's approach offers a better descriptive adequacy since it is not based on the misleading assumption that only one set of the methodological choices is right and the alternative one is erroneous. In addition, the economics of economics approach to the Reinhart-Rogoff controversy is refuted by the fact that the economists affiliated to IMF (i.e. RR and GLP from the above case studies) assumingly face similar institutional limitations and incentives. Hence, this perspective does not present a coherent point of view. Second, the constructivist approach is helpful in solving the problem of recalcitrant results, namely the findings of RR and HAP or AA and GLP seem contradictory on the grounds of the realist philosophy of economics. On the other hand, if one applies Fleck's differentiation between active and passive elements of knowledge¹³, the contradiction can be argued to be spurious. The difference in the findings is the outcome of employing various active elements of knowledge, or, in other words, various research methods. However, the passive elements of

¹³ The former term was coined to refer to those elements of knowledge that directly depend on constructed definitions, calculation methods, measurements etc. Passive elements of knowledge consist of relations among the active elements and are not directly influenced. It seems to *be real, objective and true* relation even though it is only true in relation to the active elements that it is based on (Fleck 1979, p. 10). Since the active elements of knowledge are shaped by a thought style, Fleck (1979, p. 133) clearly highlighted that *scientific observation differs when two different thought styles are involved*.

knowledge (i.e. the hypotheses) can both be true since they describe the relations between the differently constructed terms such as, for instance, “average” or “spending cuts”.

CONCLUSIONS

The case studies discussed above favor the argument that economics needs to move away from believing that the results of statistical analyses are by default correct, and have a nuanced conversation about what methods are most appropriate and in which situations. Above, the author considered the two case studies of the cliometric research that produced recalcitrant conclusions on the debt threshold hypothesis and the expansionary fiscal contraction hypothesis. Hitherto commentators have misunderstood the empirical controversies and interpreted them in terms of the misuse of quantitative techniques. On the contrary, the author argued that the research methods applied by RR with HAP, and GLP with AA (apart from the spreadsheet error, which did not influence the result in a significant way) are both supported by a number of arguments and can be said to be justified to a similar degree. The analysis of the studies on the expansionary fiscal contraction hypothesis leads to a similar conclusion, i.e. some arguments seem to be equal in weight in favor of each method of indicating when austerity at the treasury was conducted. Hence, the above case studies showed that the stylized facts arrived at by cliometricians are determined by the methodological choices. Moreover, the alternatives are justified to a similar degree.

According to Fleck’s constructivism, belonging to a thought style determines the views and presuppositions that researchers, despite being unaware of it, hold. Therefore, the cliometricians belonging to different thought styles do not find the alternative methods as equally useful, right or appealing. As Boettke et al. (2014, p. 536) argued, *scientific practice is determined by philosophical positions* (and also methodological presuppositions), *which practising scientists do not in general study with the same care that they bring to mastering their discipline*, although disparate commitments to methodology lead to obtaining contrary stylized facts. RR and HAP or AA and GLP strongly opposed each other. They chose alternative methods and obtained the results that are assumed to be contradictory. However, taking into account the fact that the findings describe relations among different passive elements of knowledge (e.g. differently defined average GDP growth, alternatively indicated years of

fiscal contractions, etc.), the views of both supporters and opponents of the hypotheses cannot be said to be erroneous since they consider differently constructed realities, i.e. their conclusions are not comparable.

Deciding whether the hypotheses are true is, taking into account the present macroeconomic environment, a hard question for economic policy-makers. Especially considering the fact that there are, to the best of the author's knowledge, no appropriate theoretical accounts of the 90% threshold hypothesis and the theoretical analysis of the expansionary fiscal contraction are also divided, the results obtained by cliometricians constitute the only available evidence. It might be fruitful to understand economic fact in a pragmatist way *as something that owes utility* (Fleck 1979, p. 72) and decide (if preferring one of the contrary points of view is unavoidable due to, for instance, the mentioned purpose of economic policy) which of the contrary hypotheses is true (or should be favored, at least) on the grounds of the utility delivered by different active elements of economic knowledge. In this case, economic policy decisions should be based on these active elements of knowledge (and implied by them, passive ones) that are useful in reaching an economic agent's goals. For instance, if a government considers reducing the CAPB by more than 1.5 percent, the conclusions of the AA approach should be considered.

A number of questions remain unanswered. First, on the grounds of the sociology of knowledge, the determinants of possessing disparate presuppositions on methodology should be considered. Second, theoretical attempts aimed at discovering the theoretical mechanisms underlying the phenomena analyzed by RR, HAP, AA, and GLP should be undertaken. Third, a robust check of the constructivist explanation of the differentiated results of cliometric research should be conducted. Despite many unsolved problems, the above case studies showed that methodological commitments determine, to a high degree, the stylized facts constructed by cliometricians, which, taking into account how influential the considered articles were, highlights the necessity of a greater dose of skepticism among economic policy-makers.

REFERENCES

- Abad, L., Khalifa, K., *What are stylized facts?*, "Journal of Economic Methodology", Vol. 22, No. 2, pp. 143-156, 2015.
- Afonso, A., Jalles, J. T., *Assessing fiscal episodes*, "Economic Modelling", Vol. 37, pp. 255-270, 2014.

- Afonso, A., Martins, L., *Monetary developments and expansionary fiscal consolidations: Evidence from the EMU*, Working Papers, Lisbon School of Economics and Managements, No. 12, 2012.
- Alesina, A., Ardagna, S., *Large changes in fiscal policy: Taxes versus spending*, NBER Working Papers, No. 15438, 2009.
- Alesina, A., Perotti, R., *Fiscal adjustments in OECD countries: Composition and macroeconomic effects*, NBER Working Paper, No. 5730, 1996.
- Alesina, A., Perotti, R., *Fiscal expansions and fiscal adjustments in OECD countries*, NBER Working Paper, No. 5214, 1995.
- Baldacci E., Gupta, S., Mulas-Grandos, C., *Debt reduction, fiscal adjustment, and growth in credit-constrained economies*, IMF Working Paper, No. 238, 2013.
- Baum, A., Checherita-Westphal, C., Rother, Ph., *Debt and growth. New evidence for the euro area*, European Central Bank Working Paper, No. 1450, 2012.
- Bell, A., Johnston, R., Jones, K., *Stylised fact or situated messiness? The diverse effects of increasing debt on national economic growth*, "Journal of Economic Geography", Vol. 15, No. 2, pp. 449-472, 2014.
- Bitar, N., Chakrabarti, A., Zeaiter, H., *Were Reinhart and Rogoff right?* "International Review of Economics & Finance", Vol. 58(C), pp. 614-620, 2018.
- Blanchard, O., *Suggestions for a new set of fiscal indicators*, OECD Economics Department Working Papers, No. 79, 1990.
- Blaug, M., Mark, B., *The methodology of economics: Or how economists explain*. Cambridge University Press, 1992.
- Boettke, P., Coyne, Ch., Leeson, P., *Earw(h)ig: I can't hear you because your ideas are old*, "Cambridge Journal of Economics", 38, pp. 531-544, 2014.
- Clemens, M., *The meaning of failed replications: A review and proposal*, Discussion Paper, No. 9000, Institute for the Study of Labor (IZA), 2015.
- Cont, R., *Empirical properties of assets returns: stylised facts and statistical issues*, "Quantitative Finance", 1(2001), pp. 223-236, 2001.
- Cooley, Th., *Frontiers of Business Cycle Research*. Princeton University Press, 3, 1995.
- Dafermos, Y. (2015), *The other half of the public debt-economic growth relationship: a note on Reinhart and Rogoff*, "European Journal of Economics and Economic Policies: Intervention", Vol. 12, No. 1, pp. 20-28, 2015.
- Dafoe, A., *Science deserves better: the imperative to share complete replication files*, *PS: Political Science&Politics*, Vol. 47, No. 1, pp. 60-66, 2014.
- De Benedictis, L., Di Maio, M., *Schools of Thought and Economists' Opinions on Economic Policy*, "Eastern Economic Journal", pp. 1-19, 2014.
- Dellepiane-Avellaneda, S., *The political power of economic ideas: The case of expansionary fiscal contractions*, *The British Journal of Politics and International Relations*, Vol. 17, Issue 3, pp. 391-418, 2014.
- Egert, B., *Public debt, economic growth and nonlinear effects: Myth or reality?*, OECD Economics Department Working Papers, No. 993, 2012.
- Egert, B., *The 90% public debt threshold: The rise and fall of a stylised fact*, *Applied Economics*, Vol. 47, Issue 34-35, pp. 3756-3770, 2015.

- Fleck, L., *Genesis and development of a scientific fact*. The University of Chicago Press, Chicago 1979.
- Friedman, M., *Essays in positive economics*. University of Chicago Press, Chicago 1953.
- Goulas, E., Zervoyianni, A., *Growth, deficits and uncertainty: Theoretical aspects and empirical evidence*, The Rimini Centre for Economic Analysis Working Paper, No. 53-13, 2013.
- Grimson, J., *Measuring research impact: not everything that can be counted counts, and not everything that counts can be counted*, *Bibliometrics. Use and Abuse in the Review of Research Performance*. Portland Press, pp. 29-41, 2014.
- Guajardo, J., Leigh, D., Pescatori, A., *Will it hurt? Macroeconomic effects of fiscal policy: evidence from industrial and developing countries*, "World Economic Outlook", pp. 93-124, 2010.
- Guajardo, J., Leigh, D., Pescatori, A., *Expansionary austerity: New international evidence*, IMF Working Paper, No. 158, 2011.
- Hamilton, J. (2013), *Reinhart-Rogoff and Herndon-Ash-Pollin: A few further comments*, Econbrowser. Analysis of current economic conditions and policy, April 2013, http://econbrowser.com/archives/2013/04/reinhartrogoff_1, access: 18 September 2015.
- Herndon, Th., Ash, M., Pollin, R., *Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff*, "Cambridge Journal of Economics", Vol. 38, Issue 2, pp. 257-279, 2014.
- Kaldor, N., *Capital accumulation and economic growth*, pp. 177-222. Palgrave Macmillan UK, 1961.
- Kersan-Skabic, I., *Public debt and interest rate spread in the European Union*, "Argumenta Oeconomica", Vol. 1, Issue 36, pp. 5-29, 2016.
- Kourtellos, A., Stengos, Th., Ming Tan, Ch., *The effect of public debt on growth in multiple regimes*, The Rimini Centre for Economic Analysis Working Paper, No. 12-60, 2012.
- Kumar, M., Woo, J., *Public debt and growth*, IMF Working Paper, No. 10/174, 2010.
- Lee, S., Park, H., Seo, M., Shin, Y., *A contribution to the Reinhart and Rogoff debate: Not 90 percent but maybe 30 percent*, CeMMAP working papers, No. 39/14, 2014.
- Maziarz, M., *Cięcia budżetowe. Przegląd najnowszych badań [Budget cuts. The overview of the latest studies]*, „Ekonomia XXI Wieku”, 2(10), pp. 9-21, 2016.
- Maziarz, M., *The Reinhart-Rogoff controversy as an instance of the 'emerging contrary result' phenomenon*, "Journal of Economic Methodology", 24(3), pp. 213-225, 2017.
- Minea, A., Parent, A., *Is high public debt always harmful to economic growth? Reinhart and Rogoff and some complex nonlinearities*, "Etudes et Documents", No. 18/2012, 2012.
- Morgan, M., *Modelling as a method of enquiry* [in:] Morgan, M., *The world in the model: How Economists work and think*, pp. 4-35. Cambridge University Press, Cambridge 2009.
- Muslu, K., Brun, Y., Meliou, A., *Preventing data errors with continuous testing*, Proceedings of the 2015 International Symposium on Software Testing and Analysis, pp. 373-384, New York 2015.
- Okalow, S., *Economist fight!*, "Canadian Business", <http://www.canadianbusiness.com/blogs-and-comment/economist-fight/>, 19 April 2013, access 4 March 2015).

- Perotti, R., *The austerity myth: Gain without pain?*, NBER Working Paper, No. 17571, 2011.
- Pescatori, A., Sandri, D., Simon, J., *Debt and growth: Is there a magic threshold?*, IMF Working Paper, No. 14/34, 2014.
- Reinhart, C., Reinhart, V., Rogoff, K., *Debt overhangs: Past and present*, NBER Working Paper, No. 18015, 2012.
- Reinhart, C., Rogoff, K., *Growth in a time of debt*, "American Economic Review", Vol. 100, Issue 2, pp. 573-578, 2010a.
- Reinhart, C., Rogoff, K., *Debt and growth revisited*, VOX. CEPRs Policy Portal, 11 August 2010, <http://www.voxeu.org/article/debt-and-growth-revisited> (access: 18 September 2015), 2010b.
- Reinhart, C., Rogoff, K., *Reinhart and Rogoff: Responding to our critics*, "The New York Times", 25th April 2013, http://www.nytimes.com/2013/04/26/opinion/reinhart-and-rogooff-responding-to-our-critics.html?pagewanted=2&_r=2 (access 19 September 2015), 2013a.
- Reinhart, C., Rogoff, K. (2013b), *Errata: Growth in a time of debt*, 5 May 2013, Harvard University, http://www.carmenreinhardt.com/user_uploads/data/36_data.pdf (access: 19th September 2015), 2013b.
- Reiss, J., *Struggling over the soul of economics: Objectivity versus expertise [in:] Experts and consensus in social science*, pp. 131-152. Springer International Publishing, 2014.
- Romer, Ch., Romer, D., *The macroeconomic effects of tax changes: Estimates based on a new measure of fiscal shocks*, NBER Working Papers, No. 13264, 2007.
- Scheuer, B., Maziarz, M., *Philosophy of economics: The constructivist and scientific realist interpretation of macroeconomics*, "Philosophy. Sociology", Vol. 30 Issue 1, pp. 8-16, 2019.
- Sewell, M., *Characterization of financial time series*, Research Note, 11/01, University College, London 2013.
- Stevenson, B., Wolfers, J., *Six ways to separate lies from statistics*, "BloombergView", <http://www.bloombergview.com/articles/2013-05-01/six-ways-to-separate-lies-from-statistics> (access: 12 September 2015), 2013.
- Ryan, P., *The path to prosperity: Restoring America's promise: Fiscal year 2012 budget resolution*, House Committee on the Budget, 2011.
- Wray, R., *Why Reinhart and Rogoff results are crap*, "EconoMonitor", <http://www.economonitor.com/lrway/2013/04/20/why-reinhart-and-rogooff-results-are-crap/>, 20 April 2013 (4.03.2015), 2013.

Received: October 2016