

Towards a political economy of statistics

Klaus Heine^{a,*} and Erich Oltmanns^b

^a*Rotterdam Institute of Law and Economics, Erasmus University Rotterdam, Rotterdam, The Netherlands*

^b*Destatis – Federal Statistical Office of Germany, Wiesbaden, Germany*

Abstract. Presently, many countries are discussing the future of official statistical data production. As a contribution to this discussion, we shall examine in this article a number of methodological aspects of a “political economy of statistics”, focussing on “statistical operationalization”, which we see as a central challenge for data production in the field of economic and social activities. In a “political economy of statistics” it is assumed that the producers and users of statistical data behave self-interested and will try to shape the statistical infrastructure to meet their personal needs, which do not necessarily coincide with the socially optimal form of data provision. As a result, individual rationality and collective rationality may fall apart and create welfare losses for society. In this contribution we therefore ask how the process of statistical data production can be organized to benefit society in total and not only specific interest groups. To that end we shall combine insights from political economy with insights from statistical operationalization.

Keywords: Data production, statistical operationalization, statistical modeling, new political economy, statistical offices

1. Introduction

There is a broad social consensus on the role of the official statistical infrastructure in a democratic society. In 1994 the United Nations stated in the *Resolution on the Fundamental Principles of Official Statistics* (and restated in 2014): “Official statistics provide an indispensable element in the information system of a democratic society, serving the Government, the economy and the public with data about the economic, demographic, social and environmental situation. To this end, official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens’ entitlement to public information.”

However, national statistical offices (NSOs) in many countries are confronted with new challenges. NSOs have to handle significant budget reductions that force them to increase efficiency and/or to reduce staff. Additionally, there is no consensus regarding the extent to which the existing statistical infrastructure fulfils the

requirements of the *Fundamental Principles of Official Statistics*, or on how it should be further developed in order to optimise its effectiveness. In fact, official statistics are currently the focus of debate in many countries (see e.g. Eltinge et al. [3], Struijs et al. [19]). For example, it is said that some of the data on offer are obsolete or even superfluous, yet at the same time new data requirements are emerging from developments in society and the economy that are not yet provided for by official statistics. The recent economic recession and the renaissance of social indicators “beyond GDP” are well-known examples of this. In addition, it is claimed that access to official statistical data is often too difficult and too expensive, if not actually impossible, which is certainly true for many people in developing countries. Or, data users in the field of science complain that they have very little influence on how official data are produced [13]. This is a point of criticism which is echoed by businesses in particular, as they feel that they are forced to spend an excessive amount of money on official statistical surveys which are of comparatively little benefit to them. Another recent challenge is the private production of massive amounts of statistical data which runs under the heading of “big data”. Here the problem is that private companies like Google or Facebook are harvesting huge amounts of data from citizens, but that those

*Corresponding author: Klaus Heine, Rotterdam Institute of Law and Economics, Erasmus School of Law, Jean Monnet Chair of Economic Analysis of European Law, Burgemeester Oudlaan 50 (J6-59), 3000DR Rotterdam, The Netherlands. Tel.: +31 10 408 2691; E-mail: heine@law.eur.nl.

companies do not automatically fall under the regime of the *Fundamental Principles of Official Statistics*. Moreover, it is by no means guaranteed that privately produced data are reliable and/or obtained according to ethical standards. The question here is whether there is any legal obligation at all to keep strict ethical standards if the data are private property of a company (to get a flavour of these problems see the debate surrounding the manipulative study by Kramer et al. [11]). In that regard, the societal debate on how to align public and private data production as well as the use and provision of those data is still in its infancy. As a result it seems necessary to lead a more fundamental and theoretically driven debate about the role of statistical infrastructure. In our view an important first step into that direction can be made by looking at the matter through the lenses of political economy and to elaborate on a “political economy of statistics”.

Looking at public statistical data production from the angle of political economy means that it is recognized that data producers and users at all levels are seen as actors who behave purposefully with regard to their individual goals. More concretely, this means that data producers and users have specific personal interests with regard to the scope and design of statistical infrastructure. While those interests are legitimate and simply reflect the diverse statistical needs of the different actors, a political economy analysis of statistical infrastructure points to the fact that individual and collective rationality may fall apart if the aggregate of the individual goals does not lead to the socially best result; in other words if the ratio between the costs and benefits of running the system is sub-optimal. If that is indeed the case it is worthwhile to analyse more deeply what the individual interests of stakeholders are and which design principles of the public statistical infrastructure should be applied to channel individual interests towards the social optimum.

This paper is divided into six sections. Following this introduction, we provide a brief analysis of the current discussions regarding official statistics. The spotlight then moves on to some basic aspects of a *new political economy* (public choice) and its relationship to statistical infrastructure. In the fourth section, we take a thorough look at the importance of the postulate of statistical operationalization when it comes to the production of statistical data. In the fifth section, which constitutes the main part of this paper, we present various fields of application for a political economy of statistics. As there are many fields of application conceivable, a choice of issues has to be made. First of all,

the most fundamental issue is discussed: the existence of official statistics. Market failure theory is applied to justify the existence of a public production of statistical data. Once the existence of independent official statistics has been approved by market failure theory, it is possible to investigate every step in the process of statistical data production from the point of view of political economy. Thereby we will put our focus on one specific step of statistical data production: statistical operationalization. The concept of statistical operationalization has been selected because in the context of political economy this concept can serve as a tool to sort out what kind of data production should be carried out by official agencies and which type of data production can remain with private producers.¹ The paper ends with a summary and a few comments on aspects for future research.

2. The ongoing debate on official statistics

Discussions are currently being held in a number of countries on the future shape of the statistical infrastructure. Notable contributions were made, for example, by Dillman [2], Fellegi [4] and Sundgren [20]. More recently the debate has become revived in a special issue of the “Journal of Official Statistics”, where Penneck [15, p. 187] summarizes the present challenges for statistical infrastructure as follows: “The redesign of statistical systems to reflect an agency-wide strategic architecture is one of the most fundamental challenges a statistics agency faces. It is both necessary and difficult. It is necessary because it will yield the cost savings and quality improvements that funders and users demand. The development of methods, systems and processes to common standards under a unified architecture yields savings in development and maintenance costs, which are timely given ever tightening budget constraints. At the same time common metadata standards, methods and approaches guarantee quality standards at a time when users become more demanding. However, although the agency-wide benefits of this approach are self-evident, statistical producers have found this approach hard to deliver.” Or, the High-Level Group for the Modernisation of Statistical Production and Services (HLG) stated 2014 in its

¹For a comprehensive review and discussion of statistical data production see also the Generic Statistical Business Process Model (GSBPM). See United Nations. Economic and Social Council. Economic Commission for Europe [22].

strategic vision: “We need to define our strategy in two directions: (a) Statistical outputs: New and better products and services more tuned to the way the world is operating today, and created from an integrated, global perspective. They will help us to stay relevant; (b) Production methods: Different and better processes and methods tuned to delivering our products at minimal cost with greater flexibility and in cooperation between institutions. This will help us to improve efficiency and effectiveness.”

These discussions have, up until now, very often bandied about arguments along the lines of a “basic supply of statistics” or “continual flow of data” which the State alone can guarantee. Similarly, the organisational principles of official statistics – particularly the obligation to provide information and the principle of confidentiality – have been evoked in order to justify the State’s involvement in the field of statistical data production. Overall, very little attention was paid to the fact that economic theory can also be employed to legitimate the public production of statistical data [6]. In particular arguments from “market failure theory” and “new political economy” can be employed. Market failure theory argues that statistical information has the characteristics of a public good and that the collection and storage of data implies economies of scale (see for example Reamer [16]). As a result the market will neither produce a sufficient amount of statistical information (public good), nor will there be a competitive pricing of the statistical information. If one takes dynamic effects into account, too, we cannot be sure how much innovation a private producer of statistical data will come up with. This is not to say that a public producer of statistical data will provide the information in an efficient way easily. But one can assume that the public production of statistical data has some generic economic advantages when compared against private production.

So far it was assumed that the individuals involved in the production of statistical data behave rational in the sense that the private production of data suffers from market failure, but that government can cure that by public production. This implies that the political process behind the public production of statistical data runs without flaws. The political decision-maker is assumed to be a benevolent agent who does not make self-interested or strategically motivated decisions, but who decides in the interest of society as a whole. However, that must not necessarily be the case. There is indeed a political process behind the production of statistical data that is driven by political interests (politi-

cians), rent-seeking groups (industry) and bureaucracy (departmental red tape). That political process cannot be assumed to deliver the desired results, since it is unconstrained. Hence, in a political economy of statistics, institutions and law play a decisive role because they are the instruments which constrain the political process and, moreover, give it its democratic legitimacy. In the following section we will seek to illustrate the specific interfaces between the production of statistical data and economic theory.

3. The production of statistical data from the perspective of new political economy

The approach of new political economy (public choice theory) is a branch of political economy theory which stresses the interdependence of the political and economic systems. It is based on the assumption that agents in the field of economic policy (politicians, economic policy advisers, bureaucrats, etc.) are primarily self-interested and do not necessarily aim at improving the total welfare of society. This approach stresses the implementation problem of policies and is different from standard welfare economics which is about the design of efficient policies and simply assumes that a benevolent legislator will introduce the policy. Thus, the normative purpose of the approach of new political economy is to propose rules on the constitutional level which constrain the subsequent policy-making process in a way that the political agents become bound to the will of society as a whole [12,23].

From what we have said it becomes intuitively clear that information plays an important role within the policy-making process. Information is needed to make reasonable political decisions but also to monitor the political process. Its significance is stressed by the fact that actors in the political and economic arena (politicians, voters, etc.) are not fully informed. The costs to emulate statistical information individually are prohibitive. Thus, actors request statistical information in the public arena. Thereby their purpose is neither to be fully informed nor to maximise the total welfare of society, but to seek individual gain.

In this respect, the statistical infrastructure has an important role to play. Information of relevance to society is provided in the form of statistical data and is judged by different stakeholders to be of varying significance depending on the benefits they derive from it. Voters, for example, are primarily interested in key economic data (e.g. unemployment rate, inflation rate),

which are conveyed to them mainly via the mass media, whereas politicians look for a significantly broader range of data to base their legislation on it, but also to use it in political competition. The statistical information required for those purposes is not just taken directly from the supply on offer from the statistical infrastructure, but is often obtained via economic policy advisors, who themselves request information from the statistical infrastructure and/or are themselves actively engaged in the production of data [6]. Statistical data can thus be characterised as a *product* that plays a crucial role in the political process, because statistical information is necessary to prepare political decision-making as well as to monitor the political process.

At the constitutional level, when a democratic basic consensus is sought in society, the production of statistical data is already an issue. That means on a basic level it is necessary to agree on objectives relating to the statistical infrastructure. As mentioned in the introduction, there is general agreement on this point (e.g. data production shall cover economic, demographic, social and environmental issues). Detailed discussions must then be held about the concrete form these objectives shall be given within the statistical infrastructure. For example, decisions are required on those areas of data production in which public activity is justified, and on those in which private production is sufficient and legitimate. When it comes to official statistics, questions regarding the way in which data production is organised also arise. Should that process be centralised, e.g. at the level of the European Union? Should a decentralised structure be chosen, e.g. at the local level of a country? Or should a mixture of the two be chosen? Finally, a basic consensus needs to be reached on the general conditions governing the behaviour of the bodies which have been selected to make up the statistical infrastructure (rules on the transmission of data, etc.).

Hence, the purpose of a political economy analysis of statistics is to disseminate proposals on achieving a coherent statistical infrastructure that best serves the interests of society. In this context, the specific characteristics of statistical data as a *product* are crucial. As a matter of fact some of the key characteristics of this product are closely linked with statistical operationalization.

4. Statistical operationalization

The concept of operationalization is derived from Bridgman [1] whose method is called operationalism

(for a critical appraisal see Hempel [7]). Within the framework of economics and social sciences it means that the concepts developed and used by theory cannot be observed and measured in a direct manner. Instead, it is necessary to use variables that can be empirically measured. And the challenge is then to construct empirically measurable variables that correspond at the best to the concepts of theory.² In principle this touches upon the fundamental problem of model specification in statistics, when a theoretical model from social sciences (for example the Capital Asset Pricing Model) is empirically applied (for an in-depth overview see Spanos [17]).

Statistical operationalization can be portrayed as three interrelated structural elements. The first, which can be described as the *horizontal aspect of operationalization in the broadest sense*, catches the variability of a theoretical concept with regard to time and content. This means that a specific operationalization may be useful for handling one (class of) problem, but may be irrelevant in other cases. Similarly, when a suitable specification is found it does not last forever, but can at some point be rendered obsolete by economic and social progress or associated changes in models and/or social values.

The second structural element can be described as the *horizontal aspect in the narrower sense of the term*: Here the idea of operationalization is based on the assumption that there is a fundamental gulf between the theoretical concept and the measurable variable which cannot be bridged – i.e. measurable variables cannot be used to make structurally identical or isomorphic copies of the respective theoretical concept, but can only produce structurally similar or homomorphic models. Therefore a statistical specification cannot be viewed as a snapshot of reality, but it involves the creative construction of models to record reality which then triggers a learning process towards a better approximation of the underlying model [14]. And, according to this (constructivist) view, even for a given and well-defined theoretical concept there may be a variety of solutions to the problem of operationalization; that means that different statistical approximations to the theoretical model may be conceivable.

The structure is completed with a component which can be called the *vertical aspect of operationalization*. If it is assumed that the construction of numerical con-

²Popular examples are the measurement of income and capital. By now, the “quality of life” is a new object of finding measurable variables.

cepts in statistics is dictated by theory, then the theory building is of great importance, because theory provides the framework within which statistical operationalization takes place. This leads to the idea that the body of economic and social theories consists of a hierarchical structure from a simple “basic model“ via “variants of the basic model“ up to “targeted theories“ for the analysis of specific issues (Giere [5]; Stegmüller [18, p. 172]). Yet this model structure has to be applied in a logically consistent manner to the numerical concepts of statistics. This means that there are not only problems of operationalization at the horizontal level between competing theories on the same level, but also at the vertical level at widely divergent levels of theory. In addition, the numerical concepts of the specific levels of theory can be understood as specific cases of numerical concepts from general levels of theory.

It should be noted that the *vertical aspect of operationalization* involves a conceptual shift with regard to data production. It emphasises the *functional* as opposed to the *institutional* perspective. In other words, attention is given not to the gulf between data producers and data users (institutional perspective), which has hitherto marked the discussions on the shape of the statistical infrastructure, but on data production as a coherent multistage process which is targeted at specific research or policy goals (functional perspective). The latter perspective will be taken for discussing the production of statistical data from the viewpoint of economic policy. We shall demonstrate that putting the focus on the vertical aspect of operationalization in conjunction with the approach of political economy can lead to important conclusions about the optimal design of the statistical infrastructure.

5. The statistical infrastructure from the perspective of economic policy: Three points of departure

5.1. The policy-making process

The idea of statistical operationalization suggests that statistical data should be produced on the basis of a scientific theory in combination with a research or policy goal. For example, a labour market theory should be developed in combination with the aim to better understand involuntary unemployment and/or to find the means to reduce involuntary unemployment. Official data production, on the other hand, is characterised by

the fact that the applied numerical concepts are determined by legislative or legal means, which guarantees their democratic legitimacy. They are basically the outcome of political agreements between stakeholders from the spheres of politics, the economy and law. Insofar one cannot easily expect an overlap between the politically negotiated official data production and the requirements of statistical operationalization.

At very general levels of theory it can be assumed that the ideas and intentions on determining numerical concepts between various stakeholders will generally correspond and cause no bigger problems. However, if the level of specification of the underlying basic theories or the stakeholders’ interest in the findings increases, it becomes more likely that the numerical concepts are determined in accordance with the ideas of those stakeholders who can exercise the greatest influence in the political decision-making process. This can explain the dissatisfaction with the statistical infrastructure quite often expressed by agents from the field of economics and other sciences. They generally belong to “lower order“ interest groups which cannot exert sufficient power in the political process. Hence the data available on the basis of the numerical concepts selected do not – or only partially – satisfy their specific requirements. That is, the numerical concepts at hand are only adequate for a limited subset of specific theories.

This raises the issue of improving the relevant decision-making mechanisms. At first sight, it looks as if there were a need for mechanisms which give disadvantaged interest groups greater influence to decide on the statistical infrastructure (right of veto, blocking minorities, etc.), e.g. producing more tailor-made data for labour market specialists. Within the paradigm of the new political economy, however, it is doubtful whether this would lead to a satisfying qualitative improvement in the official data available. Given that (self-interested) agents will find means of extending their political influence (by forming coalitions, swapping votes, etc.), this problem would *ceteris paribus* not be resolved, but merely displaced or would pop up in more sophisticated ways in other political arenas. Thus, giving more power to seemingly disadvantaged groups may only make the problem more complex, but not solve it.

At this point, it is possible to bring the *vertical aspect of operationalization* into the debate. As a starting point it is safe to say that official data should service the demands of the widest possible range of users; it is tax payers’ money that funds the production of official

data. Therefore, the policy-making process should not be targeted at the satisfaction of specific demands concerning the statistical infrastructure (which correspond with middle or upper levels of theory), but to create a stock of data at very general levels of theory, which all data users can tap into on a vertical plane to satisfy their specific thirst for knowledge. In other words, a broad and robust statistical infrastructure is needed from which specific numerical concepts can then be derived. These more specific concepts correspond to the specific needs of higher levels of theory and hence also serve the interests of specific data users.

The advantage of such a conceptualization of data provision is apparent. While a less malleable statistical infrastructure will always give advantage to those interest groups which have the factual power to determine the official statistical infrastructure, a concept of statistical infrastructure that is focused on general levels of theory cannot be captured quite that easily by interest groups. Insofar such a concept of official statistical infrastructure can also claim democratic legitimacy. Moreover, it does not foreclose the possibility that more specific needs for data can be satisfied by extensions of the statistical infrastructure by virtue of more private initiatives.

This is not the place to discuss what physical shape such a stock of data would take, although the data would need to be much more disaggregated, which would mean that the production of official data would actually even need to be expanded. The supply of such data would also require a different emphasis, with priority no longer being given to the organisation of individual surveys, but to re-designing the statistical infrastructure as an interlocking system of surveys. Nor is it easy to answer the question which political decision-making mechanisms would be best suited for a data supply of this type. On the one hand, there is certainly a case for political consensus among the stakeholders and interest groups involved. But it is also important not to overlook the implied danger, when decisions about the official statistical infrastructure are cast into the political arena. This danger would not just arise because of the wishes of users whose voices had not been heard in previous rounds of the decision-making process, but also because it might be worthwhile to express other desires which had hitherto been covered by private data suppliers. As a result rent-seeking coalitions

might emerge in which via logrolling the production of official statistical data is enlarged beyond its socially efficient amount [12, p. 104]. That means it would be easy to find coalitions for an extension of the data supply as long as all interests of data users would be covered and the data users would not have to pay for it. Against that background, the political decision-making process needs checks and balances, because it is the taxpayer who has to cover all the incurred costs of the statistical infrastructure in the end. For this reason, a more fundamental consensus on the form which the data supply should take has to be found. It is this consideration that will be examined in greater detail in the following section.

5.2. *The organisational status of statistical offices*

The question which institution or organisation might be best suited to guarantee an efficient supply of statistical data can be approached by applying two criteria. The first concerns the independence from interests of individual user groups. If regulatory power is assigned to a specific user group, this again leads to the situation portrayed above, in which certain user groups are less well served by the design of the statistical infrastructure than other groups. The second criterion is technical competence in the production of statistical data. If emphasis is put on the systemic nature of the statistical infrastructure, the relevance of regulatory power does not just lie in identifying users' needs and assessing the importance of these needs for society, but also in achieving synergies in the production of data. Thus, social and technological aspects have jointly to be taken into account.

It is evident that statistical offices should be the first to be considered for this role. They have the technical competence to adapt users' wishes to the statistical infrastructure and to disseminate proposals for the design of the system. This comprises – in terms of the postulated systemic approach – that the role of the regulatory body is not just to organise primary surveys but also to exert influence on the data products used in the context of secondary surveys. Admittedly, they do not generally fulfil the first criterion of independence. In many countries national statistical offices (NSOs) report to a ministry, the prime minister or the president. Countries with an NSO that is not an agency of the government are rare to find.³

³A remarkable exception is Mongolia. The NSO of Mongolia is

It could, however, be countered that the potential problem concerning independence is defused to a great extent by the fact that official data producers in democratic countries are bound by strict political neutrality. In so saying, dependence here is not primarily viewed as the danger of possible partisanship or even the manipulation of data in favour of certain interest groups, but it can also lie in the choice of questions used as the basis of surveys. In this sense, the problem could undoubtedly be addressed in a more appropriate fashion if official statistics were given a status similar to that of central banks in modern democracies which are free to pursue the goal of monetary stability by their own means. However, in order not to decouple statistical offices from democratic legitimacy one could think about a duty to report their activities directly to the parliament and to make the activities transparent to the public.

5.3. Opportunities and limits of a private provision of statistical data

The final issue to be investigated is the extent to which the production of statistical data can be justified as a responsibility of the state. The point of departure here will be to contemplate whether statistical data production should not actually be organised competitively in private markets, thus providing a strong incentive to find an appropriate solution to the problem of operationalization and, moreover, to let the user directly pay for a specific statistical product.

Another specific point to consider in this context is the issue of innovation in the production of statistical data. In addition to a number of other fields, such as the development of new fields of enquiry, the design (type) of survey or the processing and dissemination of anonymised microdata, operationalization is an area in which a considerable potential for innovation exists. The introduction of some kind of competition between (private) data producers might therefore be helpful to lift innovation and to provide numerical concepts that match the needs of users at the level of more specific theories.

The issue of innovation and provision of tailor-made solutions for specific users is, however, not the only

factor that has to be taken into account. Market failure theory tells us that data from the statistical infrastructure can be ascribed the characteristics of public goods, because statistical data can be consumed on a non-competing basis. That is to say that after the data have been produced they can be used by an unlimited number of users without creating additional costs. The marginal costs of production are zero and therefore the price-mechanism does not work. On the other hand the exclusion principle can be applied; in principle it is possible to limit the access of users to data. As a result, data are a good to which a property right can be assigned (exclusion), but market pricing does not work properly (non-rivalry in consumption). However, it is generally agreed that data from the statistical infrastructure should be available to all groups of society and thus exclusion from data shall not happen. This makes statistical data in the end a public good *per definitionem*. As a consequence, it cannot be expected that private producers will provide statistical data publicly. In the opposite, users will behave as free-riders and not contribute to the public good. It is therefore likely that if data provision were purely private there would be under-investment in the statistical infrastructure.

The problem of under-investment is, however, only one side of the coin. If the analysis is extended to include aspects of new political economy and the economics of information, and hence it is assumed that data users such as politicians and economic policy advisors work in their own self-interest, one would actually expect over-investment in the production of data [6]. The rationale is that according to the new political economy paradigm, interest groups will in fact attempt to occupy a leading role in opinion-forming in the political arena. For achieving that goal, statistical data will play a decisive role, because statistical data appear to lend objective support to political opinions. Once public opinion has been steered into a certain direction it is relatively easy for interest groups to draw profit from it. Politicians may adapt their political program to win the next election, bureaucrats may find statistical support to increase their budget and incumbent industries may ask for subsidies, favourable regulations or the protection from competition.

To be more concrete, this outcome of the political market is made possible by the *horizontal aspects of operationalization* (in the narrower and broader senses of the term) and can be explained by the so-called *speculative effect* that is prominent in information economics [9,10, p. 362]. The basic argument of the *speculative effect* is that one can attempt to

an independent agency under the supervision of the Parliament. See <http://unstats.un.org/unsd/dnss/cp/searchcp.aspx>. And the Statistical Office of the Netherlands now is an independent, autonomous government agency that has a stronger position after the introduction of a new Statistical law in 2004. See van de Veen [24, pp. 3–4].

guarantee the success of irreversible investment, such as expenditure on research or a political programme, by “pumping” information into the (political) market. That is linking a theoretical model (for example about involuntary unemployment) with data drawn from a numerical concept (for example the labour market statistics) and pushing then the combination of theory and data into the political arena. This then influences the preferences of consumers or voters in such a way that they are more inclined towards the product or political programme on offer than towards other programmes, which cannot come up with a similar strong operationalization between a theoretical model and a numerical concept. Yet, the damage to social welfare does not lie in the fact that a political programme is advertised by linking a theory with data. That linkage is a crucial element of political competition and it is inherent to democracies. The negative impact on social welfare is not the advertised content as such, but the fact that the production of data is extremely extensive, as all participants in the battle to shape opinion seek to monopolise public opinion for the purpose of maximising their votes. The damage is in the arms race to conquer public opinion, whereby the ammunition is made of theories linked to numerical concepts. Thus, there is the social dilemma of over-investment parallel to that of under-investment.

It is yet not possible to draw a final conclusion on the extent to which the effects of under- and over-investment cancel each other out, lead to distortions in data production or strengthen the argument of a basic market failure in the production of statistical data. Leaving those principal concerns aside it seems reasonable to pursue a dual strategy for the production of statistical data which combines market and public components. The public provision of data should be based on numerical concepts at the most general levels of theory: It is unlikely that the innovative potential at this level can be lifted by private competitors, if they cannot exploit those innovations privately, but if the benefits of those innovations accrue automatically to all data users (underinvestment). In addition, given that the data supplied at this general level are not very specific, they are not well suited to being used as political instruments aimed at the electoral market or to extract subsidies from government for an incumbent industry. It can indeed therefore be assumed that the private provision of data would result in the statistical infrastructure being under-supplied with data from the most general level of theory.

The situation is, however, different when it comes to data at higher levels of theory. In this case, it makes

sense to have diversification and competition in the supply of statistical data. First of all, users of statistical data often have very specific problems of operationalization and the official supply of data is unlikely to satisfy the varying demands of different users. Secondly, splitting the supply of statistical data would go some way to defusing the problem of the statistical infrastructure being exploited by politicians or interest groups, as the “official” data on offer can provide a solid framework and basis for competition with statistical data at higher levels of theory. That is the privately produced data at higher levels of theory have to prove their quality on a vertical plane, by connecting smoothly and consistently to the numerical concepts of the more general levels of theory.

It would be extremely costly for private suppliers of data at higher theoretical levels to gain influence over the official production of data at general levels of theory. They would face fierce opposition from other interest groups whose profits would be redistributed. Private data suppliers may also find themselves confronted with the problem of “free-riders”, since the general nature of this level of theory means that other interest groups also benefit from changes in the official production of data. Thus a single private party would hardly be inclined to provide statistical data on a general level of theory. Even if it was profitable for a coalition of interest groups to spend money on lobbying, the public good labelled “changes in the official production of data” will not be provided if one interest group is afraid that it has to bear all the costs of changing the rules but can afterwards only internalise a fraction of the political profit.

In summary, a debate is needed about the *rules of the game* to ensure that competition in statistical data production does not degenerate, but fulfils its beneficial role. The issue is not whether statistical data should *only* be provided through free competition on private markets or *only* be supplied as a purely public good, but rather which parts of the statistical infrastructure should be provided on a private basis and which would be organised by the government and how the complementarity between these two production sites becomes sustainable. To achieve this, and in view of the basic problem of statistical operationalization, it is necessary to draw not only on insights from statistics but also from that of political economy. It is an economics approach that takes systematically into account that the actors in the political arena are self-interested and that there is a need for an institutional framework to channel private interests towards the benefit of society as a whole.

6. Conclusion

Our discussion has shown that data users' dissatisfaction with the data offered by the present statistical infrastructure is closely related to the policy-making process. The institutional independence of statistical offices is seen as an important prerequisite for improving official data production and to equilibrate the horizontal and vertical dimension of the statistical operationalization problem. In addition, the problem of under or over-investment into the statistical infrastructure can provide further justification for official data production. The application of insights from new political economy gives further hints as to which parts of data production should belong to the public and which should be provided by private markets.

However, further research should still be conducted to deepen the discussion. For example, this paper did not look into the fact that within the paradigm of new political economy, official data producers may also work in their own self-interest. Another line of research could focus in more detail on the practical implications that result from the postulate of statistical operationalization. This would seem to be essential primarily because a political economy of the statistical infrastructure should not only elucidate problems in the statistical infrastructure but also make a positive contribution to its future shape in the form of specific policy recommendations. The explanations offered above can be viewed as an initial step towards such a political economy of statistics.

Acknowledgements

We would like to thank Karin Lüdemann for excellent editorial assistance as well as a reviewer and the editor Fritz Scheuren for valuable comments on an earlier draft of the paper.

References

- [1] Bridgman, *The Logic of Modern Physics*, New York, 1927.
- [2] D.A. Dillman, Why Innovation is Difficult in Government Surveys, *Journal of Official Statistics* **12** (1996), 113–124.
- [3] J.L. Eltinge, P.P. Biemer and A. Holmberg, A Potential Framework for Integration of Architecture and Methodology to Improve Statistical Production Systems, *Journal of Official Statistics* **29** (2013), 113–124.
- [4] I.P. Fellegi, Characteristics of an Effective Statistical System, *International Statistical Review* **64** (1996), 165–187.
- [5] R.N. Giere, Epistemological Roots of Scientific Knowledge. In *Induction, Probability, and Confirmation*, Minnesota Studies in the Philosophy of Science, Vol. VI. Ed. G. Maxwell and R.M. Anderson, Jr., Minneapolis: University of Minnesota Press, 1975, pp. 212–261.
- [6] K. Heine and K. Mause, Policy Advice as an Investment Problem, *Kyklos* **57** (2004), 403–428.
- [7] C.G. Hempel, A Logical Appraisal of Operationism, *Scientific Monthly* **79** (1954), 215–220.
- [8] High-Level Group for the Modernisation of Statistical Production and Services (2014), Strategic Vision of the HLG. <http://www1.unece.org/stat/platform/display/hlgbas/Strategic+vision+of+the+HLG>.
- [9] J. Hirshleifer, The Private and Social Value of Information and the Reward to Incentive Activity, *American Economic Review* **61** (1971), 561–574.
- [10] J. Hirshleifer and J.G. Riley, *The Analytics of Uncertainty*, Cambridge, 1992.
- [11] A.D.I. Kramer, J.E. Guillory and J.T. Hancock, Experimental evidence of massive scale emotional contagion through social networks, *Proceedings of the National Academy of Sciences* **111** (2014), 8788–8790.
- [12] D.C. Mueller, *Public Choice III*, Cambridge, 2003.
- [13] OECD (2013), *New Data for Understanding the Human Condition*, OECD Global Science Forum Report on Data and Research Infrastructure for the Social Sciences, Paris.
- [14] E.S. Pearson, Statistical Concepts in the Relation to Reality, *Journal of the Royal Statistical Society, Series B* **20** (1955), 204–207.
- [15] Penneck, St. Discussion on Systems and Architectures for High-Quality Statistics Production, *Journal of Official Statistics* **29** (2013), 187–192.
- [16] A. Reamer, *Putting America to Work: The Essential Role of Federal Labor Market*, Working Paper, Brookings Institution, Metropolitan Policy Program, 2010.
- [17] A. Spanos, Foundational Issues in Statistical Modeling: Statistical Model Specification and Validation, *Rationality, Markets and Morals* **2** (2011), 146–178.
- [18] W. Stegmüller, Logical Understanding and the Dynamics of Theories, *Collected Papers on Epistemology, Philosophy of Science and History of Philosophy*, Synthese Library, Dordrecht, 91, 1977, 150–176.
- [19] P. Struijs, A. Camstra, R. Renssen and B. Braaksma, Redesign of Statistics Production within an Architectural Framework: The Dutch Experience, *Journal of Official Statistics* **29** (2013), 125–145.
- [20] B. Sundgren, Making Statistical Data More Available, *International Statistical Review* **64** (1996), 23–38.
- [21] United Nations (1994/2014). *Fundamental Principles of Official Statistics*, Resolution adopted by the General Assembly on 29 January 2014. New York.
- [22] United Nations. Economic and Social Council. Economic Commission for Europe (2014). Generic Statistical Business Process Model. Prepared by the High-Level Group for the Modernization of Statistical Production and Services for the Sixty-second plenary session of the Conference of European Statisticians in Paris, 9–11 April 2014.
- [23] V. Vanberg, *Rules and Choices*, London and New York, 1994.
- [24] Veen, Gosse van der (2007). Changing Statistics Netherlands. Driving Forces for Changing Dutch Statistics. Paper presented at the Seminar on the Evolution of National Statistical Systems. Commemorative Event for the 60th Anniversary of the United Nations Statistical Commission. New York, 23 February 2007.